TEMPERATURE AND SEQUENCE DEPENDENCE OF DNA FLEXIBILITY

BY

TUNC KAYIKCIOGLU

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Biophysics and Computational Biology
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2015

Urbana, Illinois

Adviser:

Professor Taekjip Ha

Abstract

DNA fragments shorter than the persistence length display about five orders of magnitude higher flexibility than predictions of the semi-flexible polymer models based on its persistence length. Via single molecule cyclization assays, flexibility of different constructs at various conditions can be quantitatively compared. Briefly, short double stranded DNA oligos (dsDNA) of interest with single stranded overhangs are sparsely immobilized on a slide surface. Linear and circular constructs are distinguished by the extent of fluorescence resonance energy transfer between two fluorophores covalently attached at the two termini. This observation is repeated at regular time intervals after introduction of a high salt buffer that promotes looping. Rate constants can be deduced from the relaxation towards the new equilibrium state in this new condition.

We performed single molecule cyclization experiments at different temperatures to investigate the affect of temperature on the looping kinetics. Our results suggest that both looping and unlooping rates increase dramatically as temperature is increased. We then further investigated whether this effect was due to a change of the flexibility of DNA or other temperature dependent factors by measuring the bimolecular linear dimerization rate under similar conditions. Annealing and melting rates were also highly temperature dependent. In that regard, we quantified the j-factors (a.k.a. Jacobson-Stockmayer factors), which measures the effective concentration of one end of a polymer at the location of the opposite end. Our results indicated that j-factors were poorly correlated with temperature.

Next, we developed a mathematical approach to characterize a generic experiment designed to selectively enrich flexible molecules out of a dsDNA pool with different j-factors. We used this model to predict the expected outcomes for certain specific cases based on initial properties of libraries. We then tried to experimentally characterize a single molecule enrichment protocol of flexible molecules by their loop formation tendencies that had been proposed in our laboratory. Using a DNA library containing 40 000 constructs with different base sequences of the same length, we enriched rapidly looping molecules by digestion of the unlooped molecules via exonuclease treatment. We first verified that our modus operandi indeed acts as a selective filter favoring more flexible molecules by comparing the looping rates of the remaining surface bound molecules with that of the initial state. Afterwards, we

performed such a selection assay consecutively for multiple rounds by using the output of one selection as input for the next round. We performed single molecule cyclization assays on those input, intermediate and output libraries and observed an unexpected depletion of high j-factor species during this process. We also confirmed that there is not a PCR bias towards more or less flexible molecules. Our looping rate measurements hence revealed that our technique involving multiple PCR modifications rendered the library stiffer, which we attribute to undesirable by-products that arise as a result of our protocol.

Acknowledgements

It is a typical topic of the small talks, whether the subject could finally accommodate himself/herself to the new environment. Confronting this question made me realize that I actually had not noticed that there has been a change at all. I therefore feel myself obliged to thank all the Ha group members, who supported me intellectually and socially during this transition.

I especially would like to thank Digvijay Singh, who patiently addressed my questions and troubles as I was learning conducting single molecule fluorescence experiments by trial and error. The investigation of the sequence dependence were conducted as part of a collaborative project with Aakash Basu, to whom I also owe my deepest thankfulness for his advice and endless informative discussions.

Finally, I would like to thank Professor Ha for his generous material and intellectual support, without which I would not even be exposed to this field.

Contents

1 Introduction

A gentle air turbulence within the room would be enough to entangle a mass of hair into a random coil, whereas a spool of wire would not form such a structure as easily by such moderate disturbances. Underlying this different tendency to form compact bodies is the variable degree of resistance that a certain material displays against formation of a local curvature, characterized by its stiffness. Over the past few decades, significant achievements have been made to develop tools to measure elastic properties of biopolymers at microscopic scales [1]. Using one of the various chemical linkers, a stretch of the polymer can be immobilized on a surface, while the free end may be attached to a handle object through which the molecule can be subjected to various external impacts. Magnetic bead attached constructs can be subjected to magnetic fields to apply forces or torques to investigate torsional characteristics in magnetic tweezer setups. Based on the forces exerted by a refracted infrared laser beam, usage of polystyrene beads makes application of stretching forces in optical tweezer setups possible, through which force extension responses can be observed. Alternatively, this latter measurement can be also performed by attaching one of the termini to the cantilever of an atomic force microscope (AFM). An emerging trend is replication of such experiments using microfludic devices [2].

As one of the longest polymers of the nature, structural properties of deoxyribonucleic acid (DNA) has been of great scientific interest [1]. DNA consists of linear repeats of four different nucleotides, where a covalently bound chain of deoxyribose and phosphate groups constitute a repetitive helical backbone. However, in contrast to simpler polymers, DNA contains two anti-parallel helices stabilizing each other by the copious hydrogen bonds between the complementary bases. The existence of such multiple helices abolishes the means of relaxation of torsional stresses by simple rotations around one of the phosphodiester linkages, in contrast to many synthetic polymers. Combined with the hydrophobic effect driven base stacking, this structure makes double stranded DNA (dsDNA) a relatively stiff molecule.

The stiffnesses of one dimensional objects such as polymers depend on their detailed molecular structure, however, macroscopic properties of polymers can be described by simplified models that consider the polymer chain as an idealized one dimensional object. A very simplified theory considers polymers as a jagged linear structure consisting of end to end serially connected rigid subunits of short but finite length [3]. The linear elasticity of such a discrete chain, called freely jointed chain, is provided by random orientation of the rigid subunits with respect to each other. Subunits themselves are totally inelastic and inex-

tensible. N individual segments are free to rotate around any of the joints, hence establishing N degrees of freedom in solid angles. If a force F is applied to the end of the polymer, the associated potential energy is then given by,

$$W = -Fb \sum_i cos(\theta_i) \tag{1}$$

where b denotes the so called Kuhn length, which is a polymer specific length scale characterizing the rigidity. Then the partition function is,

$$\begin{aligned} Z &= \int \prod_i d\Omega_i \; e^{\frac{Fb}{k_B T} \sum_i cos(\theta_i)} = \prod_i \int d\Omega_i \; e^{\frac{Fb}{k_B T} cos(\theta_i)} \\ &= \left(\int d\Omega \; e^{\frac{Fb}{k_B T} cos(\theta)} \right)^N = \left[\frac{4\pi k_B T}{Fb} sinh\left(\frac{Fb}{k_B T} \right) \right]^N \end{aligned} \tag{2}$$

Using this partition function, the expected mean extension can be calculated by,

$$\langle l \rangle \; = \; k_B T \; \frac{\partial ln(Z)}{\partial F} = Nb \left[coth\left(\frac{Fb}{k_B T} \right) - \frac{k_B T}{Fb} \right] \tag{3}$$

However, this model is totally entropic. The tendency to restore the coiled state upon releasing the extending force is because of the fact that extension forces the ensemble into one or only a few possible configurations, whereas the coiled state can accommodate a tremendous number of degenerate states. No energies are associated with the bending, whereas real polymers resist bending. FJC model is therefore very crude and fails to approximate the behavior of real polymers unless extension is very small. An attempt to cure this inadequacy can be made by introducing forces at each nodes that is proportional to the deviation from the previous direction, i.e.

$$W = -\frac{B}{b} \sum_{n=2}^{N} \mathbf{t_i} \cdot \mathbf{t_{i-1}} - Fb \sum_{i=1}^{N} \mathbf{t_i} \cdot \hat{\mathbf{k}} \tag{4}$$

where B is the bending modulus. Referred to as the Karatky-Porod model in the literature, this model unfortunately does not have an analytical solution, but can be approximated for very small forces only. Numerical calculations can however be performed to describe its behavior. This expected extension can be experimentally measured using AFM, which indicates that this discrete theory is also too rudimentary to describe the dsDNA force extension curves unless the applied force is low [4].

An improved model can be achieved by constructing the chain out of infinitely many segments of infinitesimal length that can account for the bending, that is $b \to 0$ as the constraint $L = Nb$ is observed. The so called Worm-Like Chain model (WLC) can describe force-extension curves with a reasonable accuracy. Briefly, WLC model assumes that the polymer is a uniform chain consisting of infinitesimal subunits. The stiffness of a chain is measured by means of a length scale called persistence length (P), along which the polymer tends to follow its initial orientation. More rigorously, the correlation of the tangential vector 1 persistence length away from a certain location decays by a factor of e. Semi-flexible polymers with shorter persistence lengths are more flexible and hence can make sharper bends. The associated energy is given by,

$$E = \frac{Pk_BT}{2} \int_0^L \left| \frac{d\mathbf{u}}{ds} \right|^2 ds - F \int_0^L cos(\theta(s))\ ds \tag{5}$$

where $\mathbf{u}$ is the unit tangent vector, and integrations are parametrized along the curvilinear path of the polymer. Mean end to end extension of the polymer can be calculated from the partition function $Z = \int d\mathbf{u}\ e^{-E[\mathbf{u}(s)]/k_BT}$ as was done above, but by integrating over all paths. However, an analytical solution of the WLC force-extension equation is not available. Still, an efficient numerical approximation scheme has been developed that has a very high accuracy [5]:

$$F = \frac{k_BT}{P} \left(\frac{1}{4\ (\frac{x}{L} - 1)^2} + \sum_{i=0}^{i=7} a_i x^i \right) \tag{6}$$

where a_i are the following numerical constants: $a_0 = -0.25$, $a_1 = 1$, $a_2 = -0.516$, $a_3 = -2.737$, $a_4 = 16.075$, $a_5 = -38.076$, $a_6 = 39.499$, $a_7 = -14.177$. This prediction of the WLC model can fit the experimental force-extension data with a dramatically high accuracy. The persistence length of the DNA was determined to be about 51nm at physiological buffer conditions and did not depend on the magnitude of the applied force [5].

Despite its great success, WLC model is not without any inadequacies. An obvious problem with the WLC model, as well as the two discrete models described above, is that the chains are idealized 1D objects that can assume any shape albeit with a free energy penalty. In contrast to this infinitesimally thin string description, real polymers harbor atoms or functional groups that physically occupy space. The steric hindrance between those moieties limit the possible states to that of a a self-avoiding walk. This results in a slightly swollen structure than a Gaussian chain, corresponding to an underestimation of the hydrodynamic radius. Such effects, however, become significant only if the polymer concentration is very

high such that the subunits meet each other in their immediate surroundings during the conformational search process. Estimations for dsDNA show that this contour length limit is about $40\mu m$, which is far longer than samples what typical dsDNA experiments involve [1]. Therefore negligence of self-avoiding effects is oftentimes justifiable.

However, WLC theory suffers from a significant weakness at the opposite end, where dsDNA contour length becomes very short. Experimental evidence suggested that DNA becomes much more flexible than predicted by the theories based on the WLC model, where the authors measured the loop formation propensity of short oligos [6]. This could be demonstrated by ligating short DNA oligomers, followed by electrophoresis to quantify intra-molecular ligations that lead to formation of closed loops compared to intermolecular reactions leading to polymerization [7]. By law of mass-action, rate of chemical reactions are proportional to the concentration of the substrate species and as both ligation reactions are essentially "bimolecular" in the sense that the reaction is expected to proceed faster if the local concentration of oligo termini is elevated. Polymerization rate is determined by the number density of the oligos in the medium, whereas looping rate is proportional to the probability that one terminal can approach the other. This latter effect is measured by the Jacobson-Stockmayer factors (a.k.a. j-factors), which is the effective concentration of one end of a polymer chain at the other end's position. Other factors being equal, j-factor can be calculated by the fraction of looping events to dimerization at very low oligo number density, i.e.

$$j = \lim_{c \to 0} \frac{K_{looping}}{K_{dimerization}} \tag{7}$$

This quantity is to be interpreted as more flexible polymers having higher j-factors. As the experiments are based on DNA annealing, the loop closure also requires a match between the phases of the two sticky ends in addition to close physical proximity of the involved termini. Calculation of the expected j-factors under this constraint suggests a virtually zero loop closure probability for DNA stretches shorter than the persistence length, with an optimum looping rate at around 250bp length [8]. Surprisingly, the experimentalists observed j-factors that are more than 2 orders of magnitude higher than the expectations based on the $\sim 50\mu m$ persistence length of the DNA.

These findings also received support from studies using different methods. In a SAXS based study, where the interference between two gold nanocrystals affixed to the ends of DNA was investigated, the authors concluded that elastic rod models were not sufficient to describe end to end distance variations that they observed [9]. Yet, single-molecule fluorescence experiments emerged as an alternative way of assessing DNA flexibility [10]. By attaching

donor-acceptor fluorophore pairs at the two termini, the looping events can be observed as fluorescence resonance energy transfer (FRET) jump events. Simultaneous monitoring of a sufficiently large pool of DNA with sticky ends enables determination of the fraction of looped molecules at a certain time point. By starting the experiments in a low salt condition in which the looping events are energetically unfavored followed by triggering closed loop formation by introducing a suitable high salt buffer, looping rates can be deduced. The authors reported similarly high j-factors that support the outcomes of the ligase assays.

Despite the developments in the experimental tools, a mechanistic understanding explaining the over-flexibility of DNA at very short length scales is still not totally discovered. A view suggests that sub-critical radii of curvature could lead to local meltings in the ds-DNA, hence locally softening the polymer [11]. Under such circumstances, the uniformly distributed bending concept of WLC could give way to sharp local bends that release the bending stress on the rest of the molecule. Called kinkable worm like chain (KWLC), this model expects higher j-factors for short oligos and this increase becomes more significant as the expected number of kinks per length increases, which may explain the differences between different DNA samples. But such bubble formations constitute a slightly higher energy state, whose occupancy levels could be higher at higher temperatures, leading to increased number of kinks per length. It could perhaps be reasonable to expect an increase in j-factors as temperature is increased, which constitutes a motivation for Part 2, where the temperature effect on loopability is investigated.

Nevertheless, this interpretation is also not devoid of controversy. Such excitations would likely give rise to bending fluctuations along the DNA. SAXS experiments indicated that the major contribution to this phenomenon is likely not by bending or twisting fluctuations along the DNA [9]. The authors argued that asymmetrical distributions were not observed at the putative position of the bends. The quadratic dependence of the number of base pairs on the observed variance suggested a long-range cooperative allosteric mechanism for stretching of short DNAs, which is also inadequately described by elastic polymer models. Another approach to improve on the shortcomings of the WLC model is by introducing a softening correction term to the persistence length [12]. Long-scale bending, which can be described with reasonable accuracy by WLC, involves uncoordinated relative movements of the stacked bases and hence constitutes a Markov process. A contour length dependent choice of the persistence length can account for the non-Markovian effects. Nucleotides could be behaving as independent subunits as far as long DNA stretches are concerned, whereas short-scale bending events involve concerted planar rotations of the stacked bases. In support of this view, all atomistic molecular dynamics simulations that consider the finite

length effects reported a gradually increasing persistence length starting from 29nm [13]. Although this model can describe the experimental data more successfully, the predicted j-factors still fall short of some of the observations by about 2 orders of magnitude [10].

Another interpretation argues that the free energy required for the bending of short DNA is overestimated [14]. If DNA is considered a homogeneous isotropic 1D idealized object, closed loop formation is confronted by a large entropic cost, since the linear configuration has many more degrees of freedom to explore. However, molecular dynamics (MD) simulations suggest that the vibrational entropy of the circular DNA is much higher than linear DNA, exceeding this macroscopic entropy barrier by an order of magnitude. Equilibrium configuration can be numerically estimated by means of harmonic approximation, which suggests that the extent of thermal fluctuations in base stacking angles are very similar in circular and free dsDNA [15]. These could mean that the overall entropic cost of closed loop formation could be lower than previous estimations. On the other hand, these experimental methodologies also received some criticisms that the employed conditions might not have been adjusted properly. The apparent j-factors in ligase assays depend on the concentration of DNA ligase, excess usage of which could cause observation of disproportionately high j-factors [16]. Single-molecule experiments employ long single stranded DNA (ssDNA) overhangs that could increase the effective contour length, as well as possibility of oligo synthesis artifacts causing local deformations that favor bending [17].

This observed over-loopability of dsDNA is highly sequence dependent. A plausible explanation is that different DNA segments may have intrinsic permanent curvatures that slightly differ from each other. Indeed, the first experiments suggesting sequence dependent curvature reported anomalously slow migrating DNA on polyacrylamide gels, which becomes prominent if there are A tracts aligning with the DNA helical repeats [18]. As a huge polyelectrolyte, DNA causes a counter ion condensation to a local 1M cation concentration independent of the ambient ionic strength. NMR experiments indicated an asymmetric and sequence dependent distribution of these condensed counter ions and this was soon confirmed by x-ray crystallography, electrophoresis and simulations. Although the results show a significant correlation between the condensation asymmetry and curvature, whether there is a causation remains controversial. Another approach focuses on different preferences of nucleotide motifs for different helical forms. This difference in tendencies leads to a patched DNA structure with alternating structural isoforms, whose helical axes may slightly differ from each other. The arrangement of these junctions with respect to each other could then determine the overall conformation. The dynamic rigidity of the DNA is also highly sequence dependent. Persistence length measurements displayed as high as 4 fold difference between

very rigid GC and highly flexible AT repeats. As protein-DNA binding may induce structural deformations, binding propensities may be used as a measure of flexibility. Use of nucleosome position maps for that purpose suggested that an arrangement of AA/TT/TA dinucleotides in helical phase whereas GC out of phase promotes flexibility. Despite the ongoing efforts to build empirical models, prediction of curvature or flexibility based on sequence information alone is still a challenge.

Thus, experimental comparison of flexibility of different short dsDNA is still a field of active research. Recently, an in vitro method to selectively enrich short dsDNA with higher j-factors have been reported [19]. In this bulk assay, a DNA library containing variable sequence stretches flanked by fixed adapter sequences were incubated in a reaction mixture containing DNA ligase and lambda exonuclease. Constructs' termini contain restriction endonuclease-derived sticky ends, which allows formation of stable closed loops. Low j-factor library elements are expected to have higher exposure to digestion due to the absence of an exonuclease attack site in a cyclic DNA. Cyclic molecules are then separated from the mixture using gel chromatography and amplified via PCR. This assay provided means to establish higher j-factor libraries. However, inter-molecular ligation needs to be minimized as much as possible, which prevents usage of high DNA densities. The need to retrieve adequate quantities of DNA necessitates very high reaction volumes exceeding 2 liters. These facts motivated the attempt of reproduction of this experiment using our single molecule strategy, which is described in Part 3.

Understanding DNA flexibility is of biological importance. In eukaryotic cells, most of the cellular DNA is confined within the vesicle encapsulated nucleus and a genome can reach as long as 40m in total contour length [18]. This long DNA needs to be compacted into a cellular sub-compartment called nucleus, a spheroid roughly $5\mu m$ in diameter. This compaction is achieved by the formation of nucleosome complexes, DNA wrapped around octameric histone protein complexes. About 150bp ($\sim$50nm) DNA forms about 1.6 turns around nucleosome core particles with a core diameter of 10nm. This configuration bends DNA to much higher curvatures than would be expected given its persistence length and is highly related to the extreme bendability that DNA displays. Without the in vivo softening, WLC DNA is so stiff that estimates point to about 50atm packaging pressure in bacteriophage heads. However, nucleus is not simply a static storage space, but rather is a vault containing the precious genomic data to which access must be intermittently granted for transcription and replication. Furthermore, expression of the genes must also be regulated, which oftentimes involves looping events that brings together regulatory units that are interspersed by other stretches of DNA. Indeed, studies on lac operons of Escherichia coli revealed that in vivo

persistence length of short DNA were as low as 6.5nm [20]. Apart from the biological curiosity, unraveling mechanical properties of DNA is also of interest, which is becoming an increasingly common nanofabrication material [18].

2 Temperature Dependence of Loopability

2.1 Materials and Methods

Surface preparation

Quartz slides and coverslips were passivated by PEG coating. Briefly, a series of holes were born on quartz slides to serve as fluid inlet and outlets. Slides were thoroughly rinsed first with ultrapure water (Millipore) then with ethanol, followed by sonication in acetone for 30 minutes. After rinsing in ultrapure water, slides were sonicated in 1M KOH for 30 minutes. Thoroughly rinsed slides were burned in open flame to remove fluorescent contaminants. Surfaces were silanized in methanol (5% acetic acid and 1% N-(2 Aminoethyl) 3-Aminopropyltrimethoxysilane - UCT Specialties) for 30 minutes and rinsed afterwards thoroughly with methanol followed by a thorough rinse with ultrapure water. Each slide and coverslip were sandwiched with PEG reaction solution (16mg PEG, <0.5mg biotin derivatized PEG in 0.84% $NaHCO_3$ (w/v) - Layson Bio.) and incubated for 4 hours at room temperature. The reaction leftovers were thoroughly rinsed with ultrapure water and all surfaces were dried by nitrogen flow. Flow channels were constructed by double sided tape and instant epoxy, and assemblies were kept at -20^oC for long term storage. Shortly before the experiment, slide was brought to room temperature and treated with 20mg/ml NeutrAvidin (Thermo Fisher) in T50 (10mM Tris-HCl pH8.0, 50mM NaCl) for 1 minute.

Looping experiments

90bp long single stranded chemically synthesized DNA oligos were purchased from Integrated DNA Technologies (IDT, HPLC purified) and suspended in ultrapure water to a final concentration of $100\mu M$. Two partially complementary strands were annealed in T50 buffer at $10\mu M$ final concentration by heating up to 95^oC followed by slow cooling at room temperature for approximately 4 hours. About 30pM dilution of the annealed constructs were incubated at 50^oC for 5-10 minutes and immobilized on the surface by biotin-NeutrAvidin linkage. The unbound constructs were flushed away with T10 (10mM Tris pH8.0, 10mM NaCl). The fluorophores were monitored under house-built dual channel prism type FRET microscope in trolox solution (1mg/ml 6-hydroxy-2,5,7,8-tetramethylchroman-2-carboxylic acid, 0.8% (w/v) B-dextrose) containing 1%(v/v) gloxy (100mg/ml glucose oxidase, $91U/\mu l$ catalase [Millipore] in T50) and optionally 0.4 g/l bovine serum albumin (NEB). After initial

data collection, looping was triggered by flowing in 50μl imaging buffer additionally containing 1M NaCl, after which the system was observed at regular time intervals. Around 10 to 15 different fields of view were monitored rapidly per each time point and results were cumulated.

JJ 1-1 design used in looping assays, where bold fonts indicate single stranded complementary overhang portions and small case "t" denotes the position of the internal biotin modification on thymine:
T1: 5' - Cy3 - **CAGAATCCGT** GCTAGTACCT CAATATAGAC TCCCTCCGGT GCCGAGGCCG CTCAATTGGT CGTAGGACTA TCCTCACCTC CACCGTTTCA-3'
B1: 3' - CGATCATGGA GTTATATCTG AGGGAGGCCA CGGCTCCGGC GAGTTAACCA GCATCCTGAt AGGAGTGGAG GTGGCAAAGT **GTCTTAGGCA** - Cy5 - 5'

Dimerization experiments

Two additional HPLC purified oligonucleotides were purchased from IDT and annealed in the two alternative configurations shown below (referred to as T2B1 and T1B2, respectively). The base sequence is mainly the same as JJ1-1, except a 10bp truncation on one of the 5' end allowing to produce 90bp double stranded DNA with 10 base overhang on one end only. Around 30pM of the T2B1 construct was immobilized via biotin-NeutrAvidin linkage as described above. The baseline FRET values were measured in 1M NaCl containing imaging buffer (IB). Dimerization was triggered by flowing in 5nM T1B2 also in 1M NaCl IB, after which data is acquired at regular intervals. Due to the high background generated by free floating oligos, the spots in the Cy3 channel are obscured, preventing absolute quantification of FRET values. However, high FRET species can still be distinguished clearly in the Cy5 channel.

Modified oligo design used to observe linear polymerization, where bold fonts indicate single stranded overhang portion and small case "t" denotes the position of the biotin modification:
T2: 5' - GCTAGTACCT CAATATAGAC TCCCTCCGGT GCCGAGGCCG CTCAATTGGT CGTAGGACTA TCCTCACCTC CACCGTTTCA - 3'
B1: 3' - CGATCATGGA GTTATATCTG AGGGAGGCCA CGGCTCCGGC GAGTTAACCA GCATCCTGAt AGGAGTGGAG GTGGCAAAGT **GTCTTAGGCA** - Cy5 - 5'

T1: 5' - Cy3 - **CAGAATCCGT** GCTAGTACCT CAATATAGAC TCCCTCCGGT

GCCGAGGCCG CTCAATTGGT CGTAGGACTA TCCTCACCTC CACCGTTTCA - 3'
B2: 3' - CGATCATGGA GTTATATCTG AGGGAGGCCA CGGCTCCGGC
GAGTTAACCA GCATCCTGAT AGGAGTGGAG GTGGCAAAGT - 5'

Single molecule data analysis

Data was acquired and single molecule spots were detected using the smCamera software developed in our lab previously with 50ms exposure time. The exported time traces were processed in GNU Octave using a custom written script. Briefly, all single molecule traces with single molecule data containing Cy3 and Cy5 channel intensities are imported and all data corresponding to the same time point from different fields of view are binned. Existence of a Cy3-donor fluorophore (D) is guaranteed by the fact that spot finding is performed using an averaged image of 10 consecutive frames acquired under 532nm laser illumination at 50ms exposure. Next, noise signal with low intensity and molecules without any Cy5-acceptor(A) signal under red laser excitation (642 nm) are discarded by intensity thresholding typically set to 50 AU. This strategy is inverted for the case with Cy5 immobilized species, so that spots are detected under red laser. The looped or dimerized species are identified by their high FRET value (ϵ) assuming 0.5 as threshold and calculated according to

$$\epsilon = \frac{\frac{I_A}{I_A+I_D} - \epsilon_{DO}}{1 - \lambda - \epsilon_{DO}} \tag{8}$$

where ϵ_{DO} denotes the uncorrected FRET value of donor only species that is ideally zero and a correction is performed ($\lambda = 0.12$) for leakage into the donor channel due to imperfect dichroic splitting. The fraction of high FRET constructs passing the quality criteria is reported as the percentage of looped or dimerized molecules. The resulting time traces were fitted with an exponentially increasing function prototype using GNU Octave to determine the best fitting parameters.

The rate of looping and unlooping are Poisson processes by themselves characterized by two rate constants

$$U \underset{k_{unloop}}{\overset{k_{loop}}{\rightleftarrows}} L \tag{9}$$

Using the law of mass action,

$$\frac{dL}{dt} = k_{loop}U - k_{unloop}L = k_{loop} - (k_{loop} + k_{unloop})L$$
$$L(t=0) \approx 0 \tag{10}$$

whose solution is given by,

$$L(t) = \frac{k_{loop}}{k_{loop} + k_{unloop}} \left(1 - e^{-(k_{loop}+k_{unloop})t}\right) \tag{11}$$

A similar case applies to the linear dimerization as well, except the fact that the dimerization is a second order process.

$$M + F \underset{k_{off}}{\overset{k_{on}}{\rightleftarrows}} P \tag{12}$$

Under the assumption that free floating oligo concentration far exceeds the quantity of immobilized constructs ($F >> M$), free floating oligo concentration is roughly constant thorughout the measurement and a similar solution can be obtained

$$P(t) = \frac{k_{on}F}{k_{on}F + k_{off}} \left(1 - e^{-(k_{on}F+k_{off})t}\right) \tag{13}$$

Bulk FRET measurements

One of the annealed constructs (T2B1 or T1B2) is diluted in 1M NaCl IB used for single molecule looping experiments. A quartz cuvette is inserted into a Cary Eclipse fluorometer (Agilent) and equilibrated to the desired temperature by a water bath coupled peltier cooling system for an ample time. Other construct is added and the mixture is mixed by pipetting and fluorescence data collection was started as rapidly as possible, causing a time delay <10s. Total final volume is $200\mu l$ in which each construct is 10nM. Data is collected with 10nm slit width at four wavelength combinations $(\lambda_{ex}, \lambda_{em})$: (520nm, 570nm), (520nm, 670nm), (640nm, 570nm), (640nm, 670nm) in a cyclic manner for 60min. A Cy3 only sample was used to correct the data for donor signal leakage into acceptor signal intensity. A quantity correlated with FRET value was calculated, without any correction for detection efficiency differences. The FRET time traces were fitted by Octave to deduce the rate constants using a custom designed script.

To derive a plausible function prototype for fittings, let there be two species containing donor and acceptor fluorophores, whose concentrations are x and y, respectively. If there is a reaction happening that brings donor and acceptor pairs into close proximity ([xy] = z, such that $x + y \underset{k_2}{\overset{k_1}{\rightleftarrows}} z$), whereas the initial reaction mixture is sufficiently dilute so that energy transfer is negligible, there will be a gradually increasing FRET signal in the course of the reaction. If the interaction is governed by the law of mass action,

$$\begin{aligned} x_o &= x + z \\ y_o &= y + z \\ \frac{dz}{dt} &= k_1 xy - zk_2 = k_1(x_o - z)(y_o - z) - k_2 z \\ \frac{dz}{dt} &= k_1 z^2 - z\left(k_2 + k_1(x_o + y_o)\right) + k_1 x_o y_o \end{aligned} \tag{14}$$

which is of the following form and hence can be analytically solved

$$\begin{aligned} \frac{dz}{Az^2 - Bz + C} &= dt \\ \frac{1}{z_1 - z_2}\int \frac{1}{z - z_1} - \frac{1}{z - z_2} dz &= \frac{1}{z_1 - z_2} ln\left(\frac{z - z_1}{z - z_2}\right) = At + k \\ \frac{z - z_1}{z - z_2} &= ke^{(z_1 - z_2)At} = ke^{\sqrt{B^2 - 4AC}t} =: f(t) \\ z(t) &= \frac{z_1 - f z_2}{1 - f} \\ z(t) &= \frac{B + \sqrt{B^2 - 4AC}\frac{1+f}{1-f}}{2A} \end{aligned} \tag{15}$$

Expected FRET value of a mixture with high and very low FRET species can be calculated, if the expected FRET value of the bound state (h) and fraction of molecules in that state are known.

$$\begin{aligned} \epsilon &= \frac{n_h A_h}{n_h A_h + n_h D_h + (1 - n_h) D_l} = \frac{n_h A_h}{n_h A_h + n_h D_h + (1 - n_h)(A_h + D_h)} \\ \epsilon &= \frac{n_h A_h}{A_h + D_h} = \frac{z}{x_o} h = \frac{h}{2Ax_o}\left(B + \sqrt{B^2 - 4AC}\frac{1+f}{1-f}\right) \end{aligned} \tag{16}$$

Note the dimensionality difference between $k_1(\frac{1}{nM.s})$ and $k_2(1/s)$. For the quality of fits, we can define a new dimensionless variable $\delta = \frac{k_2}{k_1.x_o}$. Then we can adapt the general form

of the system to our problem using,

$$
\begin{aligned}
A &= k_1 \\
B &= k_2 + k_1(x_o + y_o) \\
C &= k_1 x_o y_o \\
\epsilon &= \frac{h}{2k_1 x_o}\left(k_2 + k_1(x_o + y_o) + \sqrt{(k_2 + k_1(x_o + y_o))^2 - 4k_1^2 x_o y_o}\frac{1+f}{1-f}\right) \\
\epsilon &= \frac{h}{2}\left(\delta + 1 + \frac{y_o}{x_o} + \sqrt{\delta^2 + 2\delta\left(1 + \frac{y_o}{x_o}\right) + \left(1 - \frac{y_o}{x_o}\right)^2}\frac{1+f}{1-f}\right) \\
f(t) &= ke^{\sqrt{\delta^2 + 2\delta\left(1+\frac{y_o}{x_o}\right) + \left(1-\frac{y_o}{x_o}\right)^2}\ k_1 x_o t}
\end{aligned}
\tag{17}
$$

For our case, the two species are mixed in equal stoichiometry, i.e. $x_o = y_o$

$$
\begin{aligned}
\epsilon &= \frac{h}{2}\left(\delta + 2 + \sqrt{\delta^2 + 4\delta}\frac{1+f}{1-f}\right) \\
f(t) &= ke^{\sqrt{\delta^2+4\delta}\ k_1 x_o t}
\end{aligned}
\tag{18}
$$

If the reaction is initiated by mixing the two species at $t = 0$, there should not be an appreciable FRET signal at the beginning. Due to this constraint and since x_o is under experimenter's control, there remains two degrees of freedom for the function fits on the individual traces: δ and k_1. Ideally h is also known, but we do not know in our setup, since fluorometer may have different efficiencies for detection at different wavelengths, which therefore becomes a third degree of freedom.

$$
\begin{aligned}
f(0) &= k \\
\epsilon &= \frac{h}{2}\left(\delta + 2 + \sqrt{\delta^2 + 4\delta}\frac{1+k}{1-k}\right) = 0 \\
k &= \frac{\delta + 2 + \sqrt{\delta^2 + 4\delta}}{\delta + 2 - \sqrt{\delta^2 + 4\delta}}
\end{aligned}
\tag{19}
$$

We use this k value together with Equation 18 as a function prototype to fit our FRET traces.

2.2 Experimental Results

We performed single molecule looping experiments at different temperatures by thermally stabilizing our prism type total internal reflection fluorescence (TIRF) microscopy setup via a water bath system. The sample holder, objective and the prism holder are separately served by the water circulation. All temperature measurements were recorded at the sample holder and the water bath was set to a temperature that provides the desired temperature after an ample equilibration time. We investigated the temperature range from 10^oC to 40^oC with 5^oC increments in otherwise identical conditions. Our experiments revealed a strong temperature dependence of the looping rates of the same JJ11 construct under similar initial chemical conditions (Figure 1). We observed that the relaxation rate ($k_{loop} + k_{unloop}$) towards the new equilibrium state in the high salt buffer significantly increased as temperature is increased. Direct quantification of the looping and unlooping rates was also possible by curve fitting to the exponential relaxation scheme of a two-state exponential relaxation model (Equation 11). Both of the observed looping and unlooping rates obtained by this approach had strong temperature dependence (Figure 2). Such a temperature dependence could be caused by an inherent change in the elasticity of the short oligos and is not totally counterintuitive as such a phenomenon can be observed in daily life, such as a high malleability of steel slab at high temperatures, whereas it can serve as a robust solid tool around the room temperature.

However, such an effect could also be caused by our operational definition of flexibility as loopability, which involves annealing of two complementary sticky ends at the 5' termini of our constructs. Such an annealing can only happen if two termini come in close enough proximity of each other in a permissible orientation, which involves a random 3D search of a sticky end for its counterpart purely by thermal motion. From this point of view, loopability could also be correlated with temperature due to the temperature dependence of viscosity, which is inversely proportional to the diffusion constant. The decrease in the viscosity at higher temperature could also increase the looping rates by quickening this diffusional search.

We therefore checked the temperature dependence of bimolecular association rates using an ensemble FRET approach to assess the effect of temperature dependent factors other than DNA flexibility on looping rates. We used Cy3-Cy5 labeled T1B2-T2B1 dsDNA constructs that have only sticky ends on one end to measure dimerization rates in similar buffer conditions that we used for looping rate measurements. We started the measurements with Cy3 species only, to which Cy5 labeled species was added, hence triggering a relaxation towards the dimerized state, which we can quantify by the ratio of Cy3 and Cy5 emissions (Figure 3).

The observed time dependent behavior of FRET can be described by our function prototype (Equation 19) with a high enough accuracy, which enables deduction of the dimerization rate from our measurements. By performing this experiment at different temperatures by thermally stabilizing the specimen holder of the fluorometer, we observed that dimerization rate (k_{on}) also displays a similar temperature dependence (Figure 4). Obtained disassociation rates were too low to be accurately quantified by our approach.

These preliminary findings suggested that at least some of the temperature dependence of the looping rate could perhaps be explained by the character of the annealing effect. To make a rudimentary quantitative comparison, we consider looping and unlooping as two different energy states, transition between which is subject to passing a certain activation energy barrier (ΔG^{+}). If nothing else in the experiment were temperature dependent, i.e. viscosity is constant and buffer stability is similar, we could have roughly estimated the reaction rate by adopting Arrhenius' law, where base 10 is implied:

$$\begin{aligned} k &= Ae^{\frac{-\Delta G^{+}}{k_B T}} \\ log(k) &= log(A) - \frac{\Delta G^{+}}{k_B T} log(e) \end{aligned} \tag{20}$$

If the logarithm of the reaction rate is drawn with respect to inverse of temperature, we can then approximate the free energy barrier by the slope of the best fitting line of a logarithmic plot:

$$\begin{aligned} slope &= -log(e)\frac{\Delta G^{+}}{k_B} \\ \Delta G^{+} &= -k_B \cdot ln10 \cdot slope \end{aligned} \tag{21}$$

Such a comparison of the two datasets revealed similar energy barriers (77 vs. 73 kJ/mol), indicating that the effect we observed may be attributable to energetics of dimerization rather than reflecting a real change in flexibility. Nevertheless, this does not serve as a good internal control, since in our single molecule experiments one of the reactants is immobilized on the surface by biotin-NeutrAvidin linkages, impeding the contribution of its random trajectory to the diffusion search time of the reactants. In a crude approach, we can model the overhang annealing as two reactive species diffusing in the medium that react with a certain probability if two species are within a reactive sphere of radius R_o of each other. In our setup, T1B2 and T2B1 are both of 90 bases long ssDNA molecules, therefore we can assume that they have similar diffusion coefficients through the medium (D). If both molecules are freely and

independently diffusing through the medium [21],

$$\partial_t \; p(\mathbf{r_1}, \mathbf{r_2}, t) = D(\nabla_1^2 + \nabla_2^2) \; p(\mathbf{r_1}, \mathbf{r_2}, t) \equiv \mathbf{L} \; [p] \tag{22}$$

But this operator L can be equivalently expressed in relative position $\mathbf{r} = \mathbf{r_2} - \mathbf{r_1}$ and center of mass $\mathbf{R} = \mathbf{r_1} + \mathbf{r_2}$ coordinates as $\mathbf{L} = 2D(\nabla_R^2 + \nabla_r^2)$. Then the motion along these two coordinates are separable:

$$\partial_t \; p(\mathbf{r}, t) \; P(\mathbf{R}, t) = 2D \; (\nabla_r^2 + \nabla_R^2) \; p(\mathbf{r}, t) \; P(\mathbf{R}, t) \tag{23}$$

Or this can be decomposed into two independent equations,

$$\begin{aligned} \partial_t \; p(\mathbf{r}, t) &= 2D \; \nabla_r^2 p(\mathbf{r}, t) \\ \partial_t \; P(\mathbf{R}, t) &= 2D \; \nabla_r^2 P(\mathbf{R}, t) \end{aligned} \tag{24}$$

which suggests that the relative diffusion takes place in a bulk FRET measurement with 2D. Now, assuming that inter-molecular distance distributions are isotropic, the reaction rate constant k in a bimolecular reaction is given by

$$k = \frac{4\pi R_o^2 w D}{D + R_o w} \tag{25}$$

where, R_o is the imaginary reactive radius on the surface of which reaction occurs with a rate proportional to w. Despite being a quite crude approach, this reveals that the diffusion constant significantly effects the reaction rate unless annealing reaction itself is a very inefficient process ($k_{w \to 0} \approx 4\pi R_o^2 \omega$). Since $D_{bulk} \approx 2D_{single\ molecule}$, rates obtained by bulk measurements are not a well designed comparison set with single molecule results.

We hence also designed single molecule dimerization experiments to accurately assess the k_{on} and k_{off} rate constants using the same constructs with one sticky end only. As Cy5 is a less photostable dye, we labeled the surface immobilized construct with Cy5 and detected spots under red laser to ensure that acceptor photobleaching does not lead to severe underestimations in our rate measurements. We observed that annealing curves also displayed a strong temperature dependence (Figure 5). We then deduced the k_{on} and k_{off} rate constants using a similar approach using Equation 13. As before, both k_{on} and k_{off} significantly correlated with temperature (Figure 6).

Comparing the outcomes of the two experiments, the temperature dependence mostly was

due to the temperature dependence of the inherent characteristics of our experimental setup. A more quantitative approach can be adopted by utilizing the so called Jacobson-Stockmayer factors, also known as j-factors for short. J-factor is a measure of the concentration of one termini of the polymer at the location of the other one, and is typically nonzero due to intra-molecular bending. This can be computed by the ratio of looping events to dimerizations provided that the number density of the monomers is very low, that is $j = \lim_{c \to 0} \frac{K_{looping}}{K_{dimerization}}$. Our single molecule setups can quantify these quantities separately as individual molecules are immobilized with separation greater than diffraction limit (>500nm), ruling out the possibility of any intra-molecular binding events that require a separation less than 35nm. Combining the data from the looping and dimerization experiments, we can calculate j-factors from the deduced rate constants. We resort to the indicated error inference scheme, since for a system with n technical replicas, there are n! ways of mapping the two sets and nature does not dictate any permutation over the others.

$$\begin{aligned} j &= \frac{k_{loop}\ k_{off}}{k_{unloop}\ k_{on}} = \frac{K_{loop}}{K_{dimer}} \\ \delta j &= \frac{\delta K_{loop}}{K_{dimer}} - \frac{\delta K_{dimer} K_{loop}}{K_{dimer}^2} \\ \frac{\delta j}{j} &= \frac{\delta K_{loop}}{K_{loop}} - \frac{\delta K_{dimer}}{K_{dimer}} \\ \sigma(j) &= j \sqrt{\left(\frac{\sigma(K_{loop})}{K_{loop}}\right)^2 + \left(\frac{\sigma(K_{dimer})}{K_{dimer}}\right)^2} \end{aligned} \tag{26}$$

The calculated j-factors are tabulated in Figure 7. We notice a weak correlation of j-factors and temperature, with a very low coefficient of determination ($R^2 = 0.14$). Although j-factors may vary with temperature, this variation is well within an order of magnitude, which we consider within the error margin of our experimental setup. There does not seem to be an apparent trend in the data, especially not a significant increase with the increasing temperature that is observed for macroscopic rods in our daily life.

3 Sequence Dependence of Loopability

3.1 Materials and Methods

Determination of digestion efficiency

About 30nM-50nM JJ1-1 or JJ3-3 constructs were immobilized on PEG slide surface as described before. If high substrate concentration was being tested, an additional 5nM of unlabeled JJ constructs were included, which has an identical base sequence but does not have any conjugated fluorophores. About 15-20 different fields of view (FOV) were imaged in 5mM NaCl containing imaging buffer. Imaging buffer was washed with T10, followed by infusion of digestion buffer (20U RecBCD in 50μl 1X NEB4). After 1 minute of incubation at room temperature, the sample chamber was thoroughly washed with T10 and low salt imaging buffer. This process was repeated with fresh reagents after data collection. Number of fluorophores was counted by locating spots under red and green laser illumination separately.

JJ3-3 design used in digestion efficiency quantifications, where bold fonts indicate single stranded complementary overhang portions and small case "t" denotes the position of the internal biotin modification on thymine. Internally labeled oligos were purchased from IDT as C6-amine modified thymine, which was conjugated to fluorophores via NHS ester cross-linking (courtesy of Sangwoo Park).
T3: 5' - **CAGAATCCGT** GCTAGTACCT CAATATAGAC Cy3TCCCTCCGGT GCCGAGGCCG CTCAATTGGT CGTAGGACTA TCCTCACCTC CACCGTTTCA - 3'
B3: 3' - CGATCATGGA GTTATATCTG AGGGAGGCCA CGGCTCCGGC GAGTTAACCA GCATCCTGAt AGGAGCy5TGGAG GTGGCAAAGT **GTCTTAGGCA** - 5'

Library construction

We retrieved the 'best' 10 000 Saccharomyces cerevisiae nucleosome positions from [22] and 10 000 Schizosaccharomyces pombe nucleosome positions from [23] using the unique maps that were provided among supplementary data. Generation of this subset was performed by implementing two high-pass filters on the reported SNR and positioning score values. 50 bases upstream and downstream of the selected dyad loci were retrieved from the same database the cited literature refers to (UCSC sacCer2 and PomBase v. 16.28). These 50bp

variable regions were flanked by 25bp adapter sequences at both downstream and upstream termini. This raw library was pruned to ensure at least 5 mutation spacing (deletion, insertion or substitution) between individual library components using a parallelized C^{++} Short Oligo Alignment Program (soap), that was written specifically for this project. This routine basically generates an extended library comprising each library elements and their complementary strands as different entries and pairwise aligns each different member by Needleman-Wunsch algorithm. Sequences are discarded if alignment scoring is higher than the provided threshold of 5. Four such ssDNA libraries were purchased separately as array synthesized (CustomArray). Prototype of each library element is:
5'-TTTCTTCACTTATCTCCCACCGTCC NNN...50...NNN GGCAGAAGACAAGGGAACGAAATAG-3'
where only the middle 50 bases denoted by N varies.

Obtained libraries were diluted in elution buffer (EB) of PCR purification kit (Qiagen), amplified by exponential PCR using Kapa HiFi DNA polymerase according to manufacturer's instructions ($1\mu l$ template, $2\mu M$ unmodified primers, $25\mu l$ HotStart Ready Mix in total volume $50\mu l$; 98^oC $20s, 65^oC$ $15s, 72^oC$ $6s$, with a 30s 98^oC hot start and 3min 72^oC final extension). The resulting PCR products were purified by PCR Purification kit following the manufacturer's instructions (Qiagen). This standard PCR protocol involving two primers is henceforth referred to as an exponential PCR, in connotation of the number density of the synthesized oligo which ideally grows exponentially during the process. A mixed dsDNA starting library was prepared that contains $1ng/\mu l$ of all four constituent sub-libraries containing around 10 000 sequences each, where all DNA quantifications are based on Nanodrop (Thermo).

Utilized unmodified primers, standard desalted (IDT):
5' - CTATTTCGTTCCCTTGTCTTCTGCC - 3'
5' - TTTCTTCACTTATCTCCCACCGTCC - 3'

Preparation of looping constructs

The mixed dsDNA library PCR products were amplified using 2 consecutive PCR amplifications using the same PCR protocol described above, except the fact that a 100 cycle PCR with Cy5 and biotin modified primer only was followed by a 1 cycle PCR step with the other primer. The first reaction generates many ssDNA copies by using one strand of the provided template only, whereas the second round completes this ssDNA product to dsDNA. In addition to 5' conjugated fluorophores, so obtained constructs contain overhangs at both

ends owing to an abasic site (1,2- dideoxyribose, idSp) introduced into the modified primers (purchased from IDT), across which the polymerase cannot elongate. The sample was purified both before and after the second linear amplification using PCR purification column kit. This alternative protocol provides an oligo number density ideally growing linearly with the number of reaction cycles, and therefore referred to as linear amplification in the text. The initial template material is $\sim$ 20 ng dsDNA and same as above mentioned conditions were used otherwise.

Modified primers used in this linear amplification scheme, where "t" denotes the position of the biotin modification (IDT):
5' - Cy5 - ACGGATTCTG /idSp/ CTATTTCGTTCCCTTGtCTTCTGCC
5' - Cy3 - CAGAATCCGT /idSp/ TTTCTTCACTTATCTCCCACCGTCC

Selective enrichment

About 50μl 3nM looping library was immobilized on a PEG slide using the protocol described before. To ensure that the effects of random sampling are minimized, larger than usual channel sizes were used that occupies roughly half of a full coverslip area (24mm x 50mm). The unbound molecules were washed with 200μl T10. Looping was allowed for a certain incubation time (1,3 or 5 min) in 1X NEB4 buffer diluted in water. Right after this period, unlooped molecules were digested by 2 min incubation in digestion buffer (1mM ATP, 40U RecBCD in 200μl NEB4; all purchased from NEB). Reaction was stopped by extensive buffer exchange (300 μl T10) followed by 10-15 min incubation within 10mM EDTA in formamide at 65^oC. The latter treatment releases the immobilized molecules to the chamber medium, which was collected after the incubation period to be used as template for exponential PCR amplifications. 5μl of this elute was amplified using the same exponential amplification protocol described above, typically using 12-15 cycles. To ensure sample purity, the PCR products were first column purified and then subjected to electrophoresis in 2% agarose gel (Invitrogen) and the resulting 100bp was purified out of the gel using gel extraction kit (Qiagen). The amount of retrieved material was quantified via Nanodrop and used for a further round of linear amplification.

3.2 Conceptual Understanding

We start by establishing a simple mathematical framework to draw some semi-quantitative inferences about a hypothetical experimental setup that could selectively enrich more flexible molecules out of a pool containing many different oligos. The process can be imagined as immobilization of all oligos on a surface in a condition unfavoring looping. Afterwards, a suitable action is taken that triggers looping, such as replacement of the prevalent medium by a high ionic strength buffer. This time point is to serve henceforth as the reference time point (t=0). This looping assay can be described by a relaxation to a new equilibrium state from an all-unlooped state. Following Equation 11, for each separate species within the oligo pool, the fraction of looped molecules at a given time point will be given by,

$$l(t, T, A) \equiv A(1 - e^{-t/T}) \tag{27}$$

Here $\lim_{t\to\infty} l(t, T, A) = A = \frac{k_{loop}}{k_{loop} + k_{unloop}}$ is the final equilibrium ratio and $T = \frac{1}{k_{loop} + k_{unloop}}$ is the decay time constant for this relaxation. As looping and unlooping rates by definition are non-negative both T and A are bounded, such that $T \in (0, \infty)$ and $A \in [0, 1]$.

l(t,T,A) represents the observable in single molecule looping assays, if the sample consists of single species. In case there are multiple species, the observed overall fraction of looped molecules (L) will be an average weighted by the relative amount of each construct with different relaxation parameters in the pool (ρ). That is, the observable high FRET species percentage is given by,

$$L(t) \equiv \frac{\int_0^1 \int_0^\infty \rho(T, A)\ l(t, T, A)\ dT\ dA}{\int_0^1 \int_0^\infty \rho(T, A)\ dT\ dA} \tag{28}$$

We now experimentally apply a selective filter at time t' which favors looped molecules over unlooped molecules, such as by introducing an exonuclease that preferentially digests linear molecules. We then amplify the remaining products, which ideally rescales all the oligo counts by a scalar, but may not do so in reality. If this overall process has a recovery rate ϵ for looped molecules accompanied by some random noise η due to experimental imperfections, we can quantify the remaining fraction of species after selection for each construct by,

$$s(t', T, A) \equiv \eta + \epsilon\ l(t', T, A) \tag{29}$$

Then, if another round of single molecule looping assay is performed on this new library that has been recovered from the selection, the looping rate is expected to change significantly from the original library. This is because of the fact that more flexible molecules should have been able to loop more in the allocated incubation time period before the application time point of exonuclease (t'). Hence, their representation in the product library is expected to increase at the expense of rigid molecules. For this reason, observable relaxation should drift to a new state with different kinetic variables,

$$R(t,t') \equiv \frac{\int_0^1 \int_0^\infty \rho(T,A)\ s(t',T,A)\ l(t,T,A)\ dT\ dA}{\int_0^1 \int_0^\infty \rho(T,A)\ s(t',T,A)\ dT\ dA} \tag{30}$$

Solving this equation analytically is in general not straightforward. But we can approximate the behavior of this new relaxation by the usual exponentially decaying function prototype that we adopted so far,

$$r(t,t') \equiv K(t') \left(1 - e^{-t/M(t')}\right) \approx R(t,t') \tag{31}$$

In our approximation, we demand that r(t,t') converges to the same equilibrium ratio as R,

$$\begin{gathered} \lim_{t\to\infty}\ r(t,t') = \lim_{t\to\infty}\ R(t,t') \\ K(t') = \frac{\int_0^1 \int_0^\infty \rho(T,A)\ s(t',T,A) \lim_{t\to\infty} l(t,T,A)\ dT\ dA}{\int_0^1 \int_0^\infty \rho(T,A)\ s(t',T,A)\ dT\ dA} \\ K(t') = \frac{\int_0^1 \int_0^\infty \rho(T,A)\ s(t',T,A)\ A\ dT\ dA}{\int_0^1 \int_0^\infty \rho(T,A)\ s(t',T,A)\ dT\ dA} \end{gathered} \tag{32}$$

As the second constraint for this two degrees of freedom prototype, it is reasonable to expect that the average time spent before looping is reflected accurately, that is

$$\begin{gathered} \langle t \rangle = \int_0^\infty t\ \dot{r}\ dt = \int_0^\infty t\ \dot{R}\ dt \\ \frac{K(t')}{M(t')} \int_0^\infty t\ e^{-t/M}\ dt = \frac{\int_0^1 \int_0^\infty \rho(T,A)\ s(t',T,A)\ \int_0^\infty t\ \dot{l}(t,T,A)\ dt\ dT\ dA}{\int_0^1 \int_0^\infty \rho(T,A)\ s(t',T,A)\ dT\ dA} \\ K(t')\ M(t') = \frac{\int_0^1 \int_0^\infty \rho(T,A)\ s(t',T,A)\ A\ T\ dT\ dA}{\int_0^1 \int_0^\infty \rho(T,A)\ s(t',T,A)\ dT\ dA} \end{gathered} \tag{33}$$

Or substituting for K(t') from Equation 32,

$$M(t') = \frac{\int_0^1 \int_0^\infty \rho(T,A)\ s(t',T,A)\ A\ T\ dT\ dA}{\int_0^1 \int_0^\infty \rho(T,A)\ s(t',T,A)\ A\ dT\ dA} \tag{34}$$

Note that this selection scheme can be used multiple times sequentially, in which case we could proceed in the same fashion and keep applying different selection filters, such that

$$R(t,t^{(1)},t^{(2)},...,t^{(n)}) = \frac{\int_0^1 \int_0^\infty \rho(T,A)\ \prod_{i=1}^{n} s(t^{(i)},T,A)\ l(t,T,A)\ dT\ dA}{\int_0^1 \int_0^\infty \rho(T,A)\ \prod_{i=1}^{n} s(t^{(i)},T,A)\ dT\ dA} \tag{35}$$

There is no generic analytical answer to further simplify those quantities. Nevertheless, as it is desirable to understand the expected experimental behavior of a library, for now we would like to focus our attention on a couple of specific cases.

Case 1: Homogeneous library

If the input library contains no diversity in looping characteristics, we expect that arbitrarily many rounds of selection assay under any condition chosen should not provide any contrast. In other words, there can be enrichment only if there is variability. If all the elements in the input library relax to the equilibrium level A' within relaxation time T', we can then express

$$\rho(T,A) = N\ \delta^2(T-T',A-A') \tag{36}$$

where N is an arbitrary normalization constant and δ^2 is the 2D Dirac delta function. By substituting this into Equation 35,

$$R(t, t^{(1)}, t^{(2)}, ..., t^{(n)}) = \frac{\int_0^1 \int_0^\infty N\delta^2(T - T', A - A') \prod_{i=1}^{n} s(t^{(i)}, T, A)\ l(t, T, A)\ dT\ dA}{\int_0^1 \int_0^\infty N\delta^2(T - T', A - A') \prod_{i=1}^{n} s(t^{(i)}, T, A)\ dT\ dA}$$
$$R(t, t^{(1)}, t^{(2)}, ..., t^{(n)}) = \frac{l(t, T', A') \prod_{i=1}^{n} s(t^{(i)}, T', A')}{\prod_{i=1}^{n} s(t^{(i)}, T', A')} = l(t, T', A') \tag{37}$$

which is nothing but the same relaxation equation as a single species with the respective parameters. Therefore a single species sample should behave exactly the same in looping characteristics, before and after any treatment, with any selection condition or duration.

Case 2: No selection

In this case, the assay is equally transmissible for both looped and unlooped constructs and therefore there is no contrast ($\eta \in \mathbb{R}^+ \wedge \epsilon = 0$). This could happen, for instance, if the enzyme is not able to catalyze the depolymerization reaction at all due to improper buffering conditions or its action is impeded by the molecules sticking to the surfaces. Alternatively, the time allowed for looped prior to digestion may be too short that none of the molecules can ever loop ($t' \to 0$). For the observed relaxation of the product library, this gives,

$$K = \frac{\int_0^1 \int_0^\infty \rho(T, A)\ \eta\ A\ dT\ dA}{\int_0^1 \int_0^\infty \rho(T, A)\ \eta\ dT\ dA} = \langle A \rangle \tag{38}$$

$$K\ M = \frac{\int_0^1 \int_0^\infty \rho(T, A)\ \eta\ A\ T\ dT\ dA}{\int_0^1 \int_0^\infty \rho(T, A)\ \eta\ dT\ dA} = \langle A\ T \rangle \tag{39}$$

So, the output library behaves according to the ensemble average values of the input library parameters, which is also an intuitive outcome.

Case 3: Perfect selection

Now the experimental setup works ideally, in the sense that all noise is eliminated ($\eta = 0$) and only looped molecules are recovered ($\epsilon \in \mathbb{R}^+$). So, $s(t, T, A) = \epsilon\ l(t, T, A)$, substituting

which into Equation 32 gives,

$$K(t') = \frac{\int_0^1 \int_0^\infty \rho(T,A)\ \epsilon\ l(t',T,A)\ A\ dT\ dA}{\int_0^1 \int_0^\infty \rho(T,A)\ \epsilon\ l(t',T,A)\ dT\ dA}$$
$$K(t') = \frac{\int_0^1 \int_0^\infty \rho(T,A)\ A^2\ (1-e^{-t'/T})\ dT\ dA}{\int_0^1 \int_0^\infty \rho(T,A)\ A\ (1-e^{-t'/T})\ dT\ dA} \tag{40}$$

And similarly using Equation 34, we get

$$M(t') = \frac{\int_0^1 \int_0^\infty \rho(T,A)\ A^2\ T\ (1-e^{-t'/T})\ dT\ dA}{\int_0^1 \int_0^\infty \rho(T,A)\ A^2\ (1-e^{-t'/T})\ dT\ dA} \tag{41}$$

Obtaining at least a semi-quantitative understanding of those parameters are of interest, which we can do under some simplifying assumptions. The boldest assumption is necessitated by the fact that the initial distribution of different molecules in the starting library ($\rho(A,T)$) is unknown, nor it is an easily measurable experimental quantity. Each of the 40000 sequences in the library should be ideally equally represented, but this is not the case due to biases during the chemical synthesis and PCR amplifications. Experimental determination is possible by high-throughput sequencing, but there may be deviations due to sequencing biases. Even if an accurate distribution could be obtained, their individual kinetic parameters need to be measured one at a time, making this task cumbersome. Therefore we instead assume that equilibrium and relaxation time parameters of DNA libraries of the same length are concentrated around some average values and are independent, which can be described by convolution of two Gaussian distributions,

$$\rho(A,T) = N\ exp\left(-\frac{(T-\langle T\rangle)^2}{2\sigma_T^2} - \frac{(A-\langle A\rangle)^2}{2\sigma_A^2}\right) \tag{42}$$

Due to the physical character of A and T, $\rho(A,T)$ can be nonzero only in [0,1]x(0, ∞) $\subset \mathbb{R}^2$. But we further assume that molecules tend to have very similar parameters, i.e. this distribution is sharply peaked. Under these assumptions, we can then approximate the equilibrium levels by

$$
\begin{aligned}
K(t') &= \frac{\int_0^1 \int_0^\infty N\ exp\left(-\frac{(T-\langle T\rangle)^2}{2\sigma_T^2} - \frac{(A-\langle A\rangle)^2}{2\sigma_A^2}\right)\ A^2\ (1-e^{-t'/T})\ dT\ dA}{\int_0^1 \int_0^\infty N\ exp\left(-\frac{(T-\langle T\rangle)^2}{2\sigma_T^2} - \frac{(A-\langle A\rangle)^2}{2\sigma_A^2}\right)\ A\ (1-e^{-t'/T})\ dT\ dA} \\
&\approx \frac{\int_{-\infty}^\infty \int_{-\infty}^\infty exp\left(-\frac{(T-\langle T\rangle)^2}{2\sigma_T^2} - \frac{(A-\langle A\rangle)^2}{2\sigma_A^2}\right)\ A^2\ (1-e^{-t'/T})\ dT\ dA}{\int_{-\infty}^\infty \int_{-\infty}^\infty exp\left(-\frac{(T-\langle T\rangle)^2}{2\sigma_T^2} - \frac{(A-\langle A\rangle)^2}{2\sigma_A^2}\right)\ A\ (1-e^{-t'/T})\ dT\ dA} \\
&= \frac{\int_{-\infty}^\infty exp\left(-\frac{(A-\langle A\rangle)^2}{2\sigma_A^2}\right)\ A^2\ dA}{\int_{-\infty}^\infty exp\left(-\frac{(A-\langle A\rangle)^2}{2\sigma_A^2}\right)\ A\ dA} \\
&= \frac{\sigma_A^2 + \langle A\rangle^2}{\langle A\rangle} = \langle A\rangle + \frac{\sigma_A^2}{\langle A\rangle}
\end{aligned}
\tag{43}
$$

Confirming our intuition that the more contrast exists in the input library, the more shift our selection assay will lead to. Interestingly, the equilibrium fraction does not depend on the looping time before digestion appreciably under these assumptions.

Following a similar approach for the characteristic decay time of the output library, we get

$$
\begin{aligned}
M &\approx \frac{\int_{-\infty}^\infty \int_{-\infty}^\infty exp\left(-\frac{(T-\langle T\rangle)^2}{2\sigma_T^2} - \frac{(A-\langle A\rangle)^2}{2\sigma_A^2}\right)\ A^2\ T\ (1-e^{-t'/T})\ dT\ dA}{\int_{-\infty}^\infty \int_{-\infty}^\infty exp\left(-\frac{(T-\langle T\rangle)^2}{2\sigma_T^2} - \frac{(A-\langle A\rangle)^2}{2\sigma_A^2}\right)\ A^2\ (1-e^{-t'/T})\ dT\ dA} \\
&= \frac{\int_{-\infty}^\infty exp\left(-\frac{(T-\langle T\rangle)^2}{2\sigma_T^2}\right)\ T\ (1-e^{-t'/T})\ dT}{\int_{-\infty}^\infty exp\left(-\frac{(T-\langle T\rangle)^2}{2\sigma_T^2}\right)\ (1-e^{-t'/T})\ dT}
\end{aligned}
\tag{44}
$$

By making a change of variables, we get

$$
\begin{aligned}
y &\equiv \frac{T-\langle T\rangle}{\langle T\rangle} \\
M &= \frac{\langle T\rangle \int_{-\infty}^\infty exp\left(-\frac{y^2\langle T\rangle^2}{2\sigma_T^2}\right)\ (y+1)\ (1-e^{-t'/(y+1)\langle T\rangle})\ dy}{\int_{-\infty}^\infty exp\left(-\frac{y^2\langle T\rangle^2}{2\sigma_T^2}\right)\ (1-e^{-t'/(y+1)\langle T\rangle})\ dy}
\end{aligned}
\tag{45}
$$

Now we can observe that evaluating the numerator is sufficient, since

$$
\begin{aligned}
H &\equiv \int_{-\infty}^\infty N\ exp\left(-\frac{y^2\langle T\rangle^2}{2\sigma_T^2}\right)\ (y+1)\ \left(1-e^{-t'/(y+1)\langle T\rangle}\right)\ dy \\
M &= \frac{\langle T\rangle H}{1-\langle T\rangle \frac{dH}{dt'}}
\end{aligned}
\tag{46}
$$

But this is a hard integral to analytically evaluate. We therefore now make use of our sharp-peakedness assumption to Taylor expand a few terms in the last integral around $\langle T \rangle$, equivalently $|y| << 1$.

$$1 - e^{\frac{-t'}{(y+1)\langle T \rangle}} \approx 1 - e^{-t'/\langle T \rangle} - \frac{t'}{\langle T \rangle} e^{-t'/\langle T \rangle} y - \left(\frac{t'^2}{2\langle T \rangle^2} - \frac{t'}{\langle T \rangle} \right) \; e^{-t'/\langle T \rangle} y^2 + O(y^3) \tag{47}$$

$$\begin{aligned} H &\approx \int_{-\infty}^{\infty} \left(1 - e^{-t'/\langle T \rangle} - \frac{t'}{\langle T \rangle} e^{-t'/\langle T \rangle} y - \left(\frac{t'^2}{2\langle T \rangle^2} - \frac{t'}{\langle T \rangle} \right) \; e^{-t'/\langle T \rangle} y^2 + O(y^3) \right) \\ &\qquad \cdot \; N \; exp\left(-\frac{y^2 \langle T \rangle^2}{2\sigma_T^2} \right) \; (y+1) \; dy \\ &= \int_{-\infty}^{\infty} N \; exp\left(-\frac{y^2 \langle T \rangle^2}{2\sigma_T^2} \right) \; \left(1 - e^{-t'/\langle T \rangle} - \frac{t'^2}{2\langle T \rangle^2} e^{-t'/\langle T \rangle} y^2 + O(y^4) \right) \; dy \\ &= 1 - e^{-t'/\langle T \rangle} - \frac{t'^2}{2\langle T \rangle^2} e^{-t'/\langle T \rangle} \frac{\sigma_T^2}{\langle T \rangle^2} \\ &= 1 - e^{-t'/\langle T \rangle} - \frac{t'^2 \sigma_T^2}{2\langle T \rangle^4} e^{-t'/\langle T \rangle} \end{aligned} \tag{48}$$

Using this approximate result, we can now evaluate

$$\frac{dH}{dt'} \approx \frac{e^{-t'/\langle T \rangle}}{\langle T \rangle} - \frac{\sigma_T^2}{2\langle T \rangle^4} e^{-t'/\langle T \rangle} \left(2t' - \frac{t'^2}{\langle T \rangle} \right) \tag{49}$$

Substituting these findings into the original equation yields,

$$M \approx \frac{\langle T \rangle \left(1 - e^{-t'/\langle T \rangle} - \frac{t'^2 \sigma_T^2}{2\langle T \rangle^4} e^{-t'/\langle T \rangle} \right)}{1 - e^{-t'/\langle T \rangle} + \frac{\sigma_T^2}{2\langle T \rangle^2} e^{-t'/\langle T \rangle} \left(\frac{2t'}{\langle T \rangle} - \frac{t'^2}{\langle T \rangle^2} \right)} \tag{50}$$

We could re-write this equation by expressing the cutoff time and the ensemble spread in units of $\langle T \rangle$, that is

$$M(t') \approx \frac{\left(1 - e^{-t'} - \frac{t'^2 \sigma_T^2}{2} e^{-t'} \right)}{1 - e^{-t'} + \frac{\sigma_T^2}{2} e^{-t'} \left(2t' - t'^2 \right)} \tag{51}$$

which we plot for various values of σ_T in Figure 8. We observe that in the absence of any spread ($\sigma = 0$), we recover the original relaxation time $\langle T \rangle$. For other cases, relaxation time is always shorter than the initial value, and this shift becomes more pronounced if the

diversity within the population is high. However, any contrast deteriorates if the incubation time is chosen too long and is eventually lost.

Case 4: General case

The conclusions reached above are based on a plethora of assumptions on various attributes of the starting library. Some of those characteristics have wider applicability, on which we now elaborate by relaxing some of the assumptions that we have made before. We now consider a "successful realistic experiment" in which more flexible molecules are favored as was intended despite some possible imperfections. But we consider any library characterized by an arbitrary initial looping parameter distribution and an arbitrary incubation period. In this case, if equilibrium level and relaxation rates are uncorrelated, a properly working selection assay should favor higher equilibrium ratios and shorter average looping times independent of the duration of the pre-digestion looping conditions. In more rigorous terms, $\epsilon > 0,\ \forall t' \in \mathbb{R}^+,\ \forall \rho(A,T) = \rho_A(A)\ \rho_T(T) : [0,1] \text{ x } (0,\infty) \subset \mathbb{R}^2 \rightarrow \mathbb{R}^+ \implies \Delta K \geq 0 \ \wedge\ \Delta M \leq 0.$

We first re-arrange the selection kernel as,

$$s(t') = \eta \left(1 + \frac{\epsilon}{\eta}\left(1 - e^{-t'/T}\right) A\right) = \eta\,(1 + q(T)A) \tag{52}$$

where we collectively referred to $\frac{\epsilon}{\eta}\left(1 - e^{-t'/T}\right)$ as q, which depends on T, but is independent of A for a particular experiment that fixes t' for all molecules. We can observe from this definition that $q \in [0, \frac{\epsilon}{\eta})$. For the sake of simplicity, we assume without loss of generality that the initial distribution $\rho = \rho_A \cdot \rho_T$ are independently normalized.

Referring to Equation 32, we can easily observe that if $s \equiv 1$, $K = \langle A \rangle$, which is the status quo before any selection. After the selection, we then get by substitution,

$$\begin{aligned} K &= \frac{\int_0^1 \int_0^\infty \rho(T,A)\ s(t',T,A)\ A\ dT\ dA}{\int_0^1 \int_0^\infty \rho(T,A)\ s(t',T,A)\ dT\ dA} \\ &= \frac{\int_0^1 \int_0^\infty \rho(T,A)\ \eta(1+qA)\ A\ dT\ dA}{\int_0^1 \int_0^\infty \rho(T,A)\ \eta(1+qA)\ dT\ dA} \\ &= \frac{\int_0^1 \int_0^\infty \rho(T,A)\ A\ dT\ dA + \int_0^1 \int_0^\infty \rho(T,A)\ A^2\ q\ dT\ dA}{\int_0^1 \int_0^\infty \rho(T,A)\ dT\ dA + \int_0^1 \int_0^\infty \rho(T,A)\ A\ q\ dT\ dA} \end{aligned} \tag{53}$$

$$
\begin{aligned}
K &= \frac{\langle A \rangle + \langle qA^2 \rangle}{1 \; + \; \langle qA \rangle} \\
&= \frac{\langle A \rangle + \langle qA \rangle \langle A \rangle - \langle qA \rangle \langle A \rangle + \langle qA^2 \rangle}{1 + \langle qA \rangle} \\
&= \langle A \rangle + \frac{\langle qA^2 \rangle - \langle qA \rangle \langle A \rangle}{1 + \langle qA \rangle}
\end{aligned}
\tag{54}
$$

which after some rearrangement gives

$$
\Delta K = \frac{cov(qA, A)}{1 + \langle qA \rangle} \tag{55}
$$

Note that q and A are by definition non-negative, and denominator is then also positive, being an average of positive quantities. So the direction of the equilibrium level drift only depends on the sign of the numerator, cov(qA,A). In general, the drift can be both ways, towards a higher or lower average j-factor. However, if T and A are independent,

$$
\begin{aligned}
\Delta K \;\; \propto \;\; cov(qA, A) &= \langle q \rangle_T \langle A^2 \rangle_A - \langle q \rangle_T \; \langle A \rangle_A^2 \\
&= \langle q \rangle \left(\langle A^2 \rangle - \langle A \rangle^2 \right) = \langle q \rangle \; \sigma_A^2
\end{aligned}
\tag{56}
$$

where subscripts indicate averaging over one variable only. Note that this latter term is strictly non-negative, which proves that $\Delta K \geq 0$. Following the same approach for the relaxation time shift we get the following from Equation 34,

$$
\begin{aligned}
M(q) &= \frac{\int_0^1 \int_0^\infty \rho(T, A) \; A \; T \; \eta \; (1 + qA) \; dT \; dA}{\int_0^1 \int_0^\infty \rho(T, A) \; A \; \eta \; (1 + qA) \; dT \; dA} \\
&= \frac{\int_0^1 \int_0^\infty \rho(T, A) \; A \; T \; dT \; dA \; + \; \int_0^1 \int_0^\infty \rho(T, A) \; q(T) \; A^2 \; T \; dT \; dA}{\int_0^1 \int_0^\infty \rho(T, A) \; A \; dT \; dA \; + \; \int_0^1 \int_0^\infty \rho(T, A) \; q(T) \; A^2 \; dT \; dA} \\
&= \frac{\langle AT \rangle + \langle qA^2 T \rangle}{\langle A \rangle \; + \; \langle qA^2 \rangle} \\
&= \frac{\langle AT \rangle}{\langle A \rangle} + \frac{\langle qA^2 T \rangle \langle A \rangle - \langle qA^2 \rangle \langle AT \rangle}{\langle A \rangle \left(\langle A \rangle + \langle qA^2 \rangle \right)}
\end{aligned}
\tag{57}
$$

using this result and noting that $M(q = 0) = \frac{\langle AT \rangle}{\langle A \rangle}$ in accordance with equation 34, we reach after some further rearrangement

$$
\Delta M = \frac{cov(qA^2 \langle A \rangle - \langle qA^2 \rangle A, T)}{\langle A \rangle \left(\langle A \rangle + \langle qA^2 \rangle \right)} \tag{58}
$$

As before, we note that the denominator is an average over product and sum of non-negative quantities and hence is positive definite, thus the sign of the shift depends on the covariance in the numerator. Once again, in the case of independence this simplifies to

$$\begin{aligned} \Delta M \ &\propto \ \langle A\rangle_A\langle A^2\rangle_A\langle qT\rangle_T - \langle T\rangle_T\langle q\rangle_T\langle A^2\rangle_A\langle A\rangle_A \\ = \langle A\rangle \ \langle A^2\rangle \ cov(q,T) \ &= \ \frac{\epsilon}{\eta}\langle A\rangle \ \langle A^2\rangle \ cov(1-e^{-t'/T},T) \\ &\propto \ -\langle A^2\rangle \ \langle A\rangle \ cov(e^{-t'/T},T) \end{aligned} \tag{59}$$

Now we observe that this term is non-positive, since $\langle A^2\rangle \geq 0$, $\langle A\rangle \geq 0$ and the given exponential function is an increasing function of T, making the covariance positive. We therefore conclude that $\Delta M \leq 0$.

3.3 Experimental Results

The conclusions arrived at above can be qualitatively understood as a population drift driven by artificial selection. In a pool of various DNA sequences, if flexible sequences have selective advantage over the others, the population should gradually shift towards more flexible sequences. Experimentally, this corresponds to a drift towards higher equilibrium levels as well as shorter relaxation times to equilibrium, as selection is repeated over and over if there is not a significant negative correlation. However, just as natural selection requires a mechanism generating heritable variants of a trait, the expected shifts critically depend on the existence of an initial variance within the input library. In fact, contrast has an intrinsic limit regardless of the allocated looping time for a given variability.

To test this idea experimentally, we measured looping times of different libraries that we subjected to sequential selections by immobilizing DNA libraries with overhangs under low salt conditions. We then replaced this medium with NEB4 buffer, which is high in salt content and hence triggers looping (50mM potassium acetate, 20mM Tris acetate, 10mM Magnesium acetate, 1mM DTT). After a brief incubation time, we digested the unlooped molecules by RecBCD (a.k.a. exonuclease V). We hypothesize that looped molecules are immune to digestion as there are no free entry sites for the digestion to be initiated by an exonuclease, provided there is no or negligible non-specific endonuclease activity of the enzyme. We indeed confirmed this by subjecting the JJ1-1 construct described above to looping conditions in NEB4 for 20min after immobilization, followed by digestion buffer that

contains RecBCD as well as the imaging buffer so that fraction of looped molecules can be quantified during the digestion process (Figure 9). Owing to the very small amount of molecules immobilized ($\sim$ 1 billion), we observed a very rapid selective digestion of the linear molecules, as assessed by the rapid jump of the percentage of high FRET species due to the introduction of the exonuclease at the 21st minute of incubation.

In single molecule assays, neighboring fluorophores need to be separated from each other by at least as large as the diffraction limit, whereas selection assays can work at far higher molecule densities, as long as the formation of large circular multimeric structures by joining of multiple molecules is avoided. The optical distinguishability requires a separation as large as at the micron scale, whereas the latter condition can be satisfied provided that mean separation between molecules is greater than the contour length of the oligos, which is around 35nm. Because recovery of as many molecules as possible after selection is desirable to minimize the effects of random drifts, achieving high enough densities of immobilized molecules is desirable. For this reason, we mixed labeled and unlabeled JJ1-1 oligos and repeated the digestion experiment, which confirm that digestion of unlooped molecules is essentially complete within minutes (Figure 10). Due to the very low number of oligos per chamber volume, 2 min digestion time was sufficient even when the immobilized construct density reached 5nM.

As exonuclease digestion starts from the termini, digestion may be far from complete even though the fluorophores conjugated to 5' termini are released rapidly. To verify that the digestion proceeds to a point at least as far as adapter sites within the allocated incubation time, we replicated this digestion test with internally labeled JJ3-3 constructs. This condition is necessary so that only the selected molecules will be able to bind primers during the downstream amplification procedures. The fluorophores in this case are away from the termini by 24 and 30 bases, whose presence we assessed independently by counting the spots under green and red laser illumination separately. We observed that most of the molecules were digested beyond the adapter sites, although fraction of remaining molecules was higher as well as the digestion speed after 2 minutes of incubation (Figure 11). We further confirmed that this depletion of low FRET species is not caused by an unpredicted photophysical effect of the buffer components that RecBCD requires for efficient catalytic activity (Figure 12).

Next, we applied this idea to a library containing 40.000 different sequences that was chemically synthesized. Since the observable changes in the looping kinetics critically depend on the diversity of the input sample, we pooled sequences that originate from budding and fission yeast. We introduced the same structure with biotin as anchors and 10bp sticky ends

by means of PCRs with modified primers. In the nth cycle of an exponential PCR, the products formed at the (n-1)th cycle are melted, most of which then bind the free primers and are completed to a dsDNA product. However, this process is not 100% efficient and the unelongated molecules may re-anneal to each other after the temperature is cooled down. As the library consists of molecules whose terminal adapter regions are complementary, but the middle variable segments are not, there is a risk that this process may lead to formation of permanently bent molecules due to bubbling. We therefore used a linear amplification scheme, in which only one strand is synthesized as ssDNA and then completed to dsDNA in a 1 cycle reaction only. Using this strategy provides almost perfectly unlooped constructs, which are properly loopable in high salt buffer (Figure 13). Out of a starting material of 20ng, typical yield is about $1\mu g$, which puts the overall efficiency of this whole process to 50%.

We tested our selective enrichment on this library by immobilizing 100nM library along with 5nM unlabeled JJ constructs to mimic high sample density conditions. We then allowed looping in imaging buffer with 1X NEB4 buffer for 5 minutes, after which the chamber was incubated with RecBCD for 2 minutes. The reaction was terminated by an extensive wash with 10mM EDTA in T10, and molecules were brought back to the initial state by incubation in low salt imaging buffer for an hour. We then measured the looping rates of these remaining molecules in NEB4 and compared with the relaxation kinetics of the original library (Figure 14). We indeed observed an increased looping rate, especially at early times, hence validating our selective enrichment premise.

We then used this strategy to generate new libraries by amplification of the remaining molecules by PCR. Briefly, we incubated immobilized products on the PEG passivated surface in NEB4 for 5 minutes after which digestion buffer was applied for 2 minutes. The remaining constructs were eluted and amplified using exponential PCR amplification with unmodified primers. The resulting dsDNA library was used as template for 3 consecutive rounds of sticky end generation and selection. The incubation time in the looping buffer was reduced from 5min to 3min and 1 min for the consecutive rounds to increase the selective pressure (Figure 15). We then measured and compared the looping rate of the input and selection products in NEB4 buffer (Figure 16). As a surprise, the libraries became more and more rigid as the selective pressure was increased as evidenced by the steeply decreasing j-factors. We also observed that this situation is not caused by a side-effect of the NEB4 buffer components, but also is the case in our simple imaging buffer containing 1M NaCl only (Figure 17).

A possible explanation could be that the PCR conditions were somehow favoring stiffer templates more compared to the flexible strands. Although the number of cycles in each exponential amplification step is limited to as few as 15 cycles only, any possible bias could have drastic effects on the resulting library due to the accumulation of those effects throughout the enrichment procedure. We therefore tested this possibility by subjecting our library to a further 60 cycles of PCR, in addition to that required for the initial 15 cycle amplification necessary for the sample preparation. However, we observed that the extended numbers of PCR cycles lead to unexplainable side products with very high molecular weights, which we attribute to the occasional recombination events and self-priming of the accidentally matching sequences due to the high complexity of the template library. Another concern is that the PCR reaction does not proceed in an exponential trend, but rather tend to reach a saturation level after a certain period due to the depletion of the reaction substrates such as primers. This typical sigmoidal amplification trend is undesirable, if small biases are to be made observable. We therefore split the 60 cycle amplification into 15 cycle sub-reactions. After each reaction, the reaction mixture was column purified and the amount of the product was quantified. About 0.2ng of the product was used as template in the subsequent round that is set afresh using new reagents.

Unlike the over-amplification case that leads to extraneous bands, this procedure did not lead to any observable product other than 100bp in agarose gel electrophoresis (Figure 18). We then subjected the initial and final libraries to the linear amplification scheme described above and compared the looping rates of these products (Figure 19). We observed a slight increase in the looping relaxation rate after extended cycles of PCR, albeit this effect was very small. We also observed that the looped population fraction after the equilibrium is slightly decreased. However, compared to the dramatic drop that we observed in our sequential selection assays, this shift after 60 cycles was nothing but minor. As a comparison, the final sample in the selection trial (S3) has 45 more PCR cycles than the starting material and the observed equilibrium drop is as high as 50%. The effect of PCR bias alone therefore cannot explain this observed change. An alternative possibility could be that our amplifications or template retrieval with the modified primers might not have been working as smoothly as we imagined, but rather forming different structures that we previously were not aware of. As a check against such a possibility, we repeated the same multi-step amplification assay, but now also performing the linear amplification steps that introduce the overhangs (Figure 20). After three consecutive rounds of looping construct preparation but without any selective pressure, we did not observe a significant change in the overall looping characteristics of our library, in accordance with our expectations.

These observations suggested that our immobilization and exonuclease treatment could somehow give rise to artifacts that render a portion of our library unloopable, which we tested on a single species sample that was commercially column synthesized. As our setup uses chemically synthesized non-genomic DNA and extensive PCR amplifications with standard nucleotides, DNA flexibility that we measure should only depend on the primary base sequence but not on loss of any potential epigenetic modifications. This therefore suggests that looping curve of a specimen solely consisting of a single species should reproduce the same looping curve of the initial sample after arbitrarily many selections in any condition, which we tested following the same selection protocol. Unexpectedly, we observed that this was also not satisfied in our setup (Figure 21). Indeed, we observed that the input sample looped significantly slower than our nucleosome libraries. On the other hand, the product generated after one round of selection became significantly more flexible, but we afterwards observed a gradual decay of this flexible trend, similar to what we observed before.

Finally, accidental synthesis of unloopable molecules that carry both fluorophores as by-products could lead to this decreasingly loopable library trend. If they are somehow indigestible or very favorably replicated in PCR, they could accumulate and dominate the characteristics of the whole sample after each amplification. However, such an interpretation falls short of explaining the increased flexibility of column synthesized product that we observed. Thus, we repeated the extended PCR amplification test on this construct and compared the flexibility of the initial sample to that after 60 total additional PCR cycles (Figure 22). As before, all of the initial, intermediate or final products were purely 100bp clean products as much as the sensitivity of agarose gel electrophoresis can provide. Astonishingly, the sample displayed very high loopability after over-amplification (Figure 23), in agreement with the sudden increase we observed when we applied sequential selection assay.

4 Discussion and Future Directions

Our experiments demonstrate the the cyclizability of short DNA oligos is highly temperature dependent. As would be expected from the rudimentary Arrhenius' theory, both looping and unlooping rates significantly increase as temperature is increased. The fundamental assumption underlying cyclization experiments is that closed loop formation propensity is a good measure of how flexible a certain oligo is. However, comparison between two constructs is meaningful in this respect only if all other conditions are fixed. That is because absolute measure of loopability also heavily depends on variables other than the intrinsic flexibility of the polymer, such as viscosity of the medium and the stabilizing free energy of the overhang annealings. Indeed, in the temperature range that we studied, viscosity of the water varies by about two fold (1307 vs 653 $\mu Pa \cdot s$ at 10^{o}C vs 40^{o}C [24]).

To account for these effects, we measured the dimerization rates in a similar setup. The association and disassociation rates also demonstrated a similar temperature dependence behavior. Although kinetic parameters steeply varied with temperature, the looping rates did not get selectively enhanced over dimerization appreciably. Some of the macroscopic objects that we encounter in our daily lives tend to become more and more flexible as their temperature is increased. An example of this could be an iron slab that is highly malleable at high temperatures whereas it is very stiff at room temperature. This leads us to conclude that unlike the macro-scale materials such as a polyethylene, whose stiffness values monotonically decrease in the temperature range we studied by an order of magnitude [25], DNA does not display such a behavior. This is in contradiction to our expectations if extreme bendability is mediated by the formation of transient kinks or bubbles, whose incidence should be highly correlated with temperature.

Nevertheless, our technique has some inherent limitations. One of the most intriguing difficulties is in the stabilization of the temperature. We achieved temperature stabilization by coupling our TIRF microscope setup to a water bath whose temperature is maintained at a constant temperature. Although the temperature readings were taken from the sample holder rather than the water bath, different parts of the quartz imaging slide are differentially exposed to the ambiance and hence slight temperature variations are inevitable. Another complication is related to the temperature associated changes of the imaging medium. To avoid photobleaching, our measurements needed to be performed in an oxygen scavenger system that involves two coupled reactions of glucose oxidase and catalase. As with many enzymatic reactions, the catalyzed reaction rates are highly temperature dependent [26].

Although temperature also changes the solubility and the diffusion rate of the oxygen, our back of the envelope numerical calculations indicate that the regions of the imaging chamber reasonably away from the holes are not to exposed free oxygen significantly (not shown). But, probably the most significant impact of this on our measurements is by slightly more rapid decreases in the pH at elevated temperatures as opposed to lower. Indeed, at very high temperatures, we observe an aberrant decrease in the population of the looped state, which could be attributable to such side effects beyond our control. A potential way to overcome, or at least rule out, this effect could be by means of a flow experiment where the buffer is stored at room temperatures and slowly flowed in to the sample chamber continuously throughout the measurement process. We also used a Tris based buffering system, whose pKa is also temperature dependent, though this pH change does not exceed 0.5 in the temperature range we studied [27]. Thus, investigation of the pH could also be an interesting follow-up study. Furthermore, although photochemical events associated with temperature are not expected to affect our cyclization assay as alternating laser excitation was employed, nature of the linear dimerization measurements are unsuitable for this technique. We tried to minimize the impact of this phenomenon by labeling the immobilized constructs by Cy5, which is more labile, and free floating molecules by Cy3. The polymerization rates are therefore prone to slight underestimations, which can become significant if Cy3 photobleaching time scales are much faster than the association kinetics.

We also studied a methodology that could provide information about the sequence dependence of DNA flexibility. If successful, our enrichment process for more flexible molecules could provide a very useful tool compatible with high throughput sequencing suited for studies focused on characterization of the sequence dependence in various contexts. The numerical scheme that we developed agrees with our intuition that the libraries can be drifted towards distributions with higher j-factors and/or relaxation rates. We demonstrated that our method successfully enriches looped molecules at the expense of linear molecules in a mixed library by observing the abundance of high and low FRET species throughout the process. We also confirmed that there is a detectable contrast between output libraries obtained by such a strategy and the input library.

However, for reasons that so far evade our understanding, multitude of the chemical processes involved in our proposed protocol appears impermissive to generate daughter libraries that preserves this established contrast. The overall apparent stiffening which reveals itself as gradually decreasing equilibrium levels could normally mean a selection based on the contrast in unlooping rates rather than looping, as Equation 55 suggests. Yet, the observation of a similar trend in single species input library cases rather suggest an unpredicted instability

in our procedure, instead. A potential problem could be related to the purity of the initial samples. Most common mutations in chemical syntheses are truncation products, which should have been eliminated by purifications and PCR amplifications [28]. However, the capping reagents employed in the syntheses are not 100% effective, and deletion mutations are therefore inevitable. Any bias towards such products can lead to a change in the segment length distribution, hence can lead to detectable changes in the flexibility due to changes in the relative phasing of the sticky ends. Another possible source of error is our employment of non-standard PCR techniques. Our methodology relies on fluorophore and dideoxyribonucleotide modified DNA oligos as primers. The existence of abasic sites, accompanied with the single stranded species within the reaction mixture could perhaps be leading to poorly characterized interactions and hence arousal of by-products. In fact, we observe that the amplification of template with overhangs are not in general robust, which we temporarily cope with by the reduction of PCR cycles and optimization of purification protocols. Although we did not observe any such effects after multiple rounds of re-amplification of templates with such sticky ends, the cross-contamination from successive reaction steps in real setups could be interfering with the quality of the downstream processes, and thus leading to artifacts that we could not yet fully characterize.

These observations may potentially mean that our current dideoxyribonucleotide approach may not be the best choice possible, but our optimization efforts will be ongoing. There are still some potential workarounds in our toolbox, including introduction of overhangs by endonuclease digestion or pruning of sticky ends before any amplification. Unless DNA based lifeforms cease to exist, there is still more to do to understand the DNA, and we are still far from reaching the limit of what single molecule techniques can provide.

5 Figures and Tables

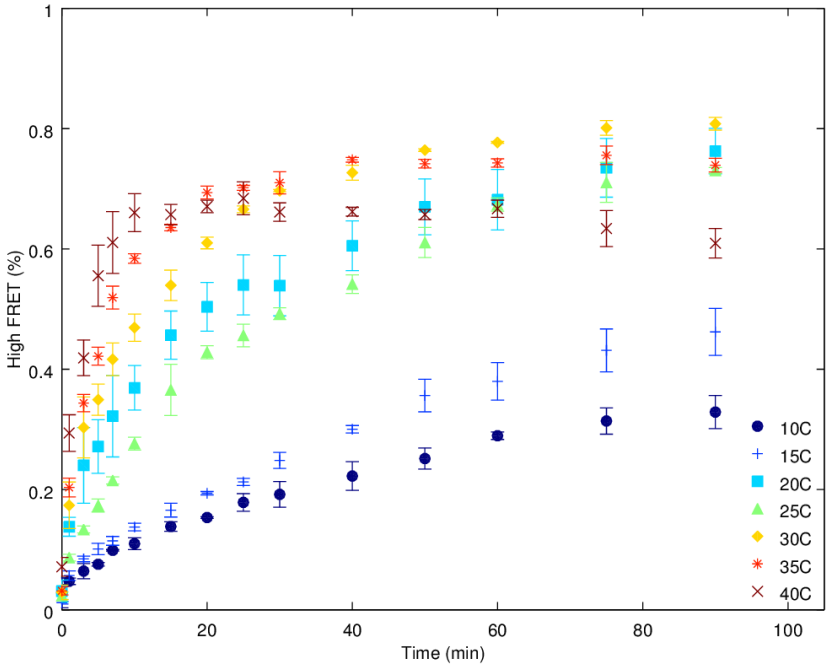

Figure 1: Looping of JJ1-1 sequences display temperature dependence. Data points are based on cumulation of 10-15 different field of views that were scanned per each time point. Binned data are reported as $\mu \pm \sigma$.

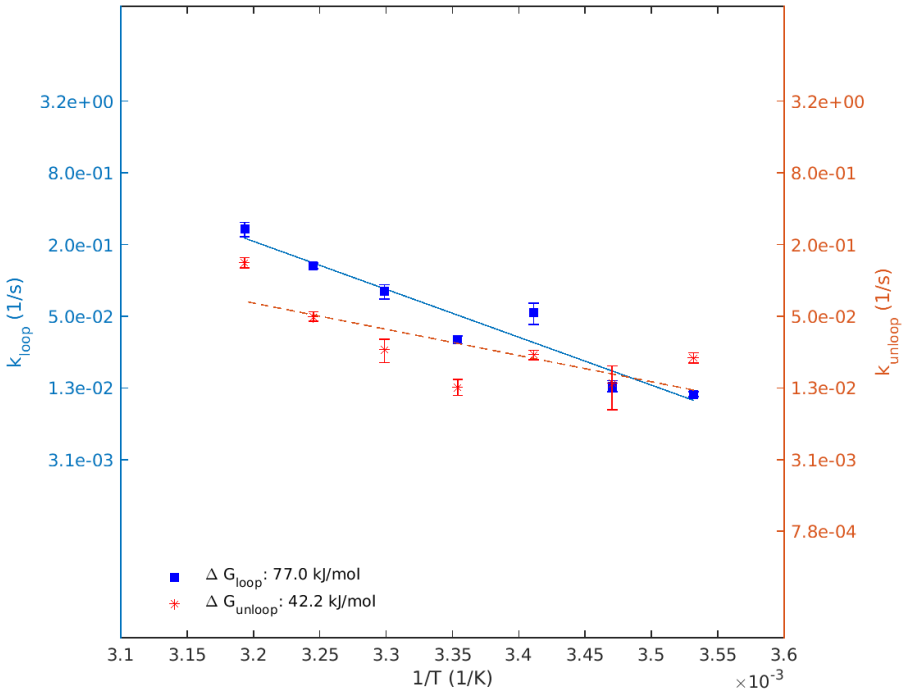

Figure 2: Looping and unlooping rates display strong temperature dependence. Both reactions are faster at higher temperatures due to the increased energy to overcome activation energy barrier as well as decreased viscosity of the medium. The slope of the trend lines reflect the expected free energy barrier predicted by Arrhenius' theory, if the temperature dependence of the other experimental conditions were ignored for looping (solid) or unlooping (dashed) transitions. Note that both vertical axes are in logarithmic scale.

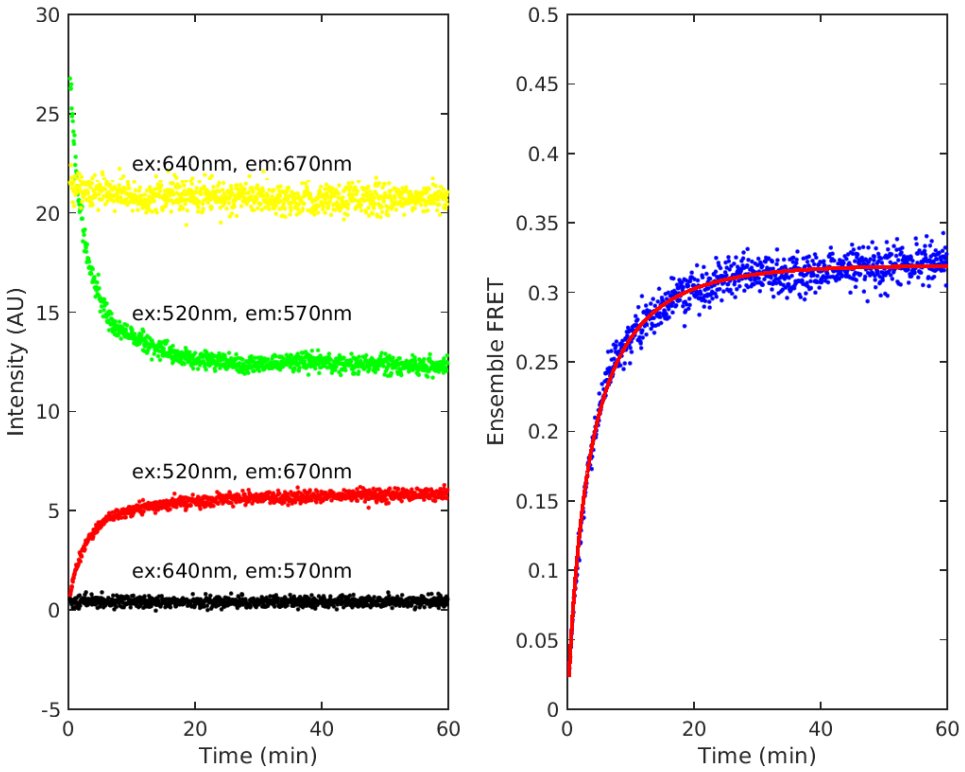

Figure 3: Dimerization of T1B2 and T2B1 can be assessed by bulk FRET measurements. Example data point, captured at 20^oC, shows that the observed increasing trend in FRET can be fitted by the function prototype derived above with high accuracy to deduce reaction rates.

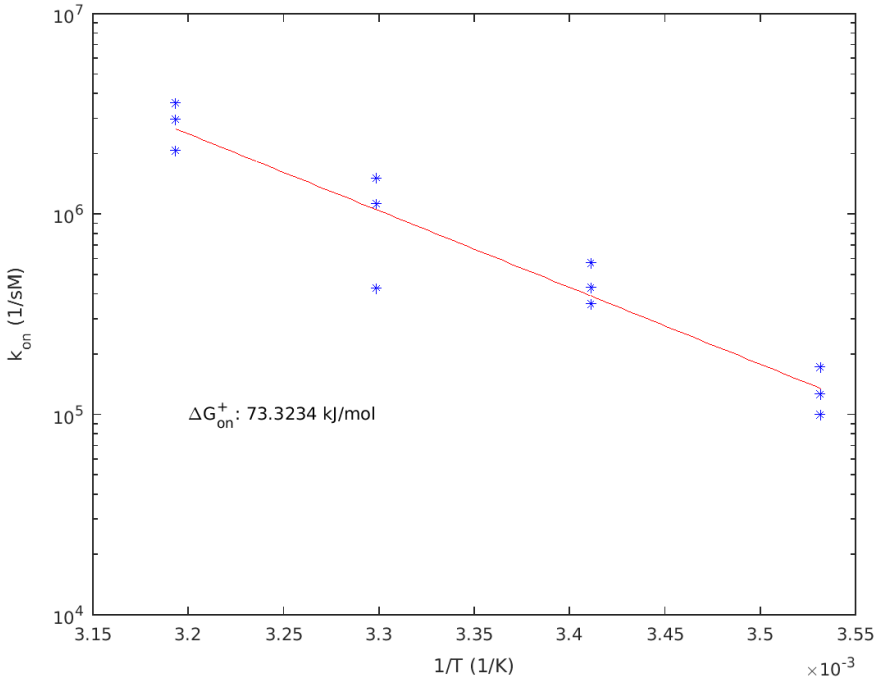

Figure 4: Dimerization rates measured by bulk FRET display strong temperature dependence. Dimerization rate k_{on} as measured by ensemble FRET at different temperatures can be reasonably fit to the equation from whose slope the free energy barrier can be approximated. This indicates a strong temperature dependence. Disassociation rate k_{off} as measured by ensemble FRET at different temperatures cannot be determined very accurately due to relative slowness of the reaction.

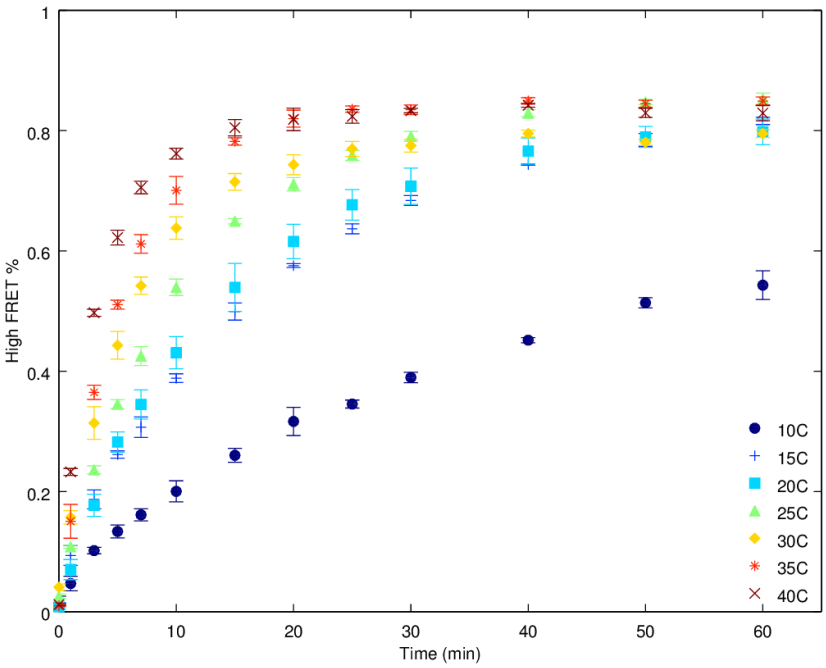

Figure 5: Dimerization of T1B2 and T2B1 also display strong temperature dependence. Data points are based on cumulation of 10-15 different field of views that were scanned per each time point. Binned data are reported as $\mu \pm \sigma$.

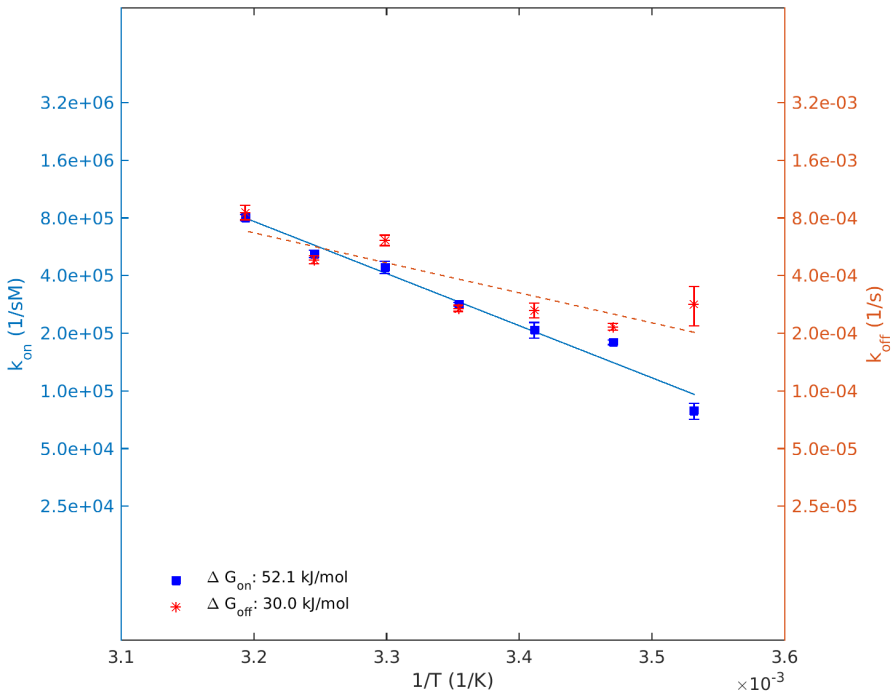

Figure 6: k_{on} and k_{off} are temperature dependent. Using the previously described fitting approach, the indicated ΔG values reflect the expected energy barrier if the annealing (solid) and melting (dashed) processes are governed by a simple two state mechanism and all other temperature dependent parameters such as viscosity contribute negligibly. Note that both vertical axes are in logarithmic scale.

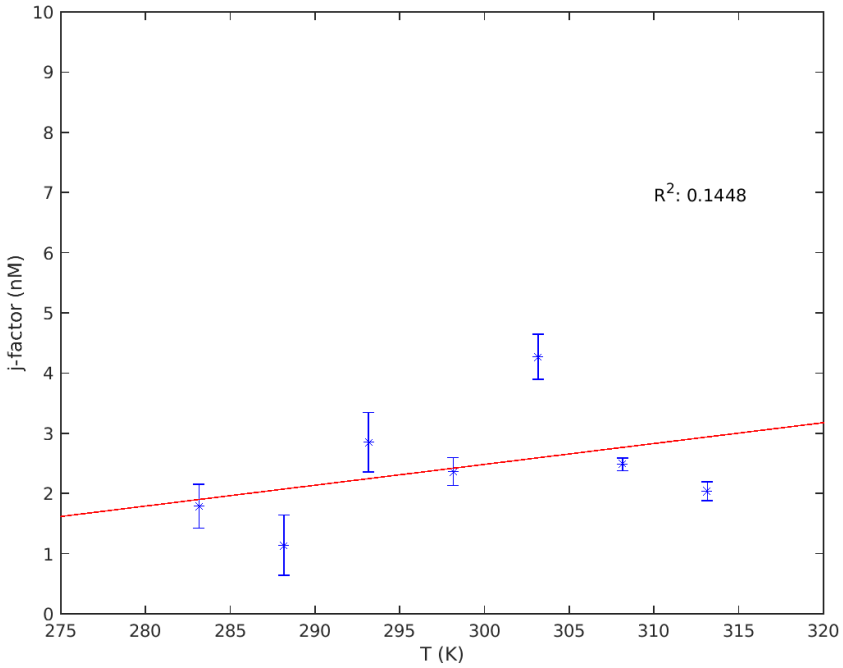

Figure 7: Measured j-factors do not depend on temperature appreciably. Although there is some correlation with the temperature, this correlation is relatively weak and coefficient of determination is very low.

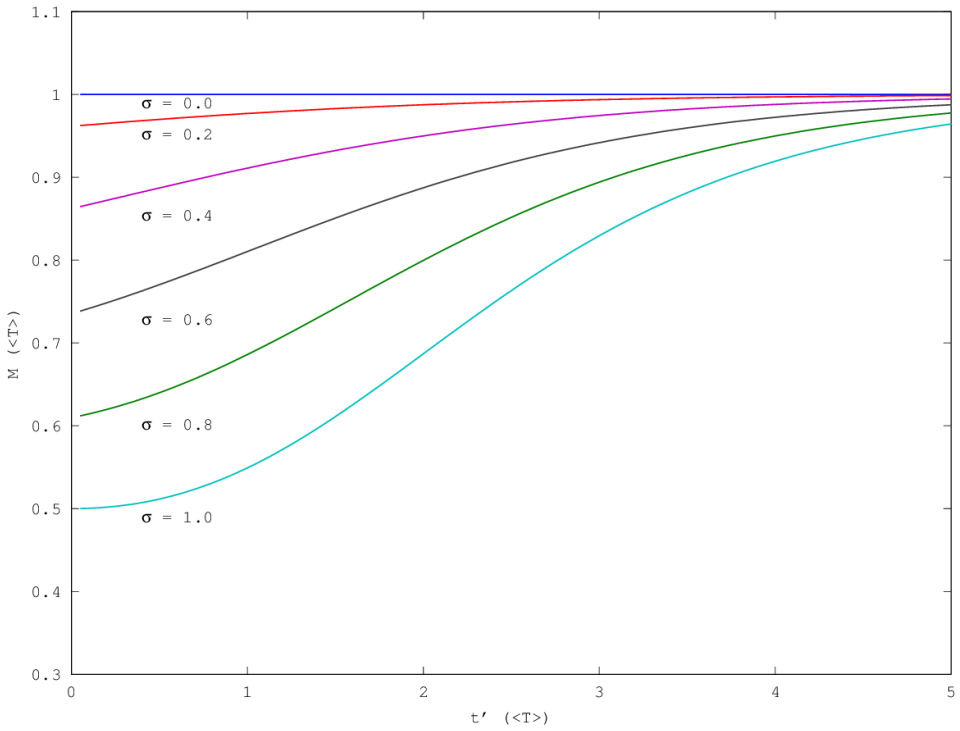

Figure 8: Theoretically estimated looping time of the output library in a selection assay if relaxation rate and equilibrium constants are independent. All time parameters are in fractions of the input library's average looping time constant $\langle T \rangle$. As the spread of the time constants in the input library becomes larger, the contrast between the input and output increases significantly. If the looping cutoff times are chosen to be very large, all cases converge to zero contrast, demonstrating the importance of the careful choice of selection times.

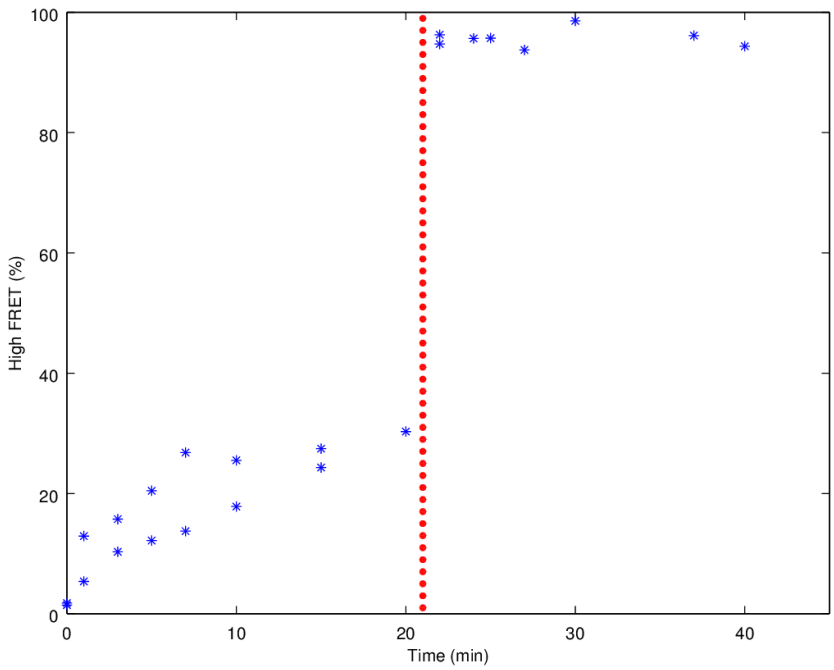

Figure 9: Exonuclease digestion is a selective filter on the looped molecules. The molecules were immobilized and imaged in imaging buffer containing 5mM NaCl. Shortly after replacement of the buffer with NEB4, the looping events were triggered (t=0min). Introduction of exonuclease causes a rapid elimination of the unlooped low-FRET molecules, characterized by a sudden jump in the fraction of high FRET molecules (t=21min).

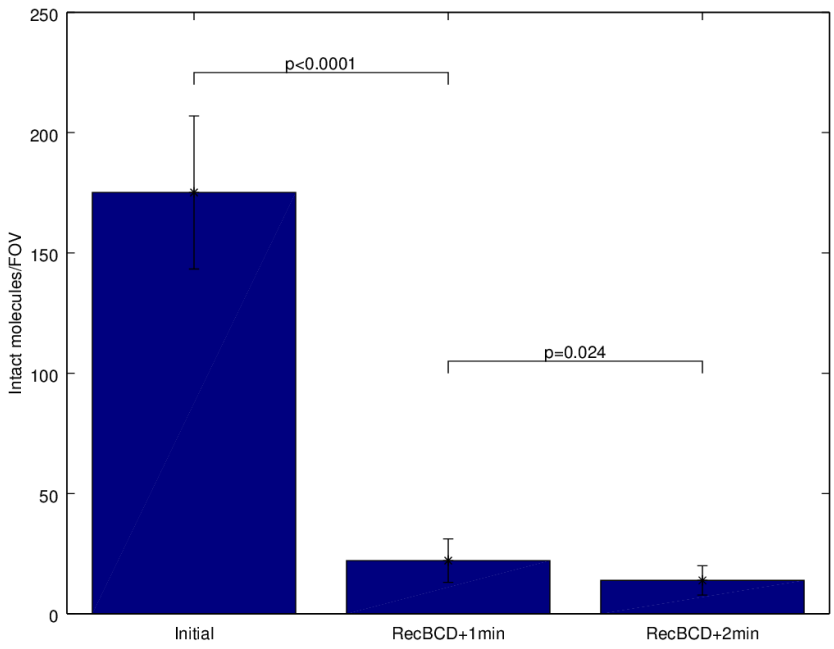

Figure 10: Fast selective digestion phenomenon is not disrupted at high substrate concentration. Immobilized JJ DNA solution concentration was as high as 5nM, which was directly subjected to digestion without any separate time allocation to allow looping. To ensure optical distinguishability, only a small fraction (<1%) of the loaded sample is fluorescently labeled. Number density of the immobilized molecules was determined via total number of spots remaining on the surface after each 1 minute incubation with exonuclease. Sample comparisons were performed by two-tailed z-test, error bars represent 1 SD.

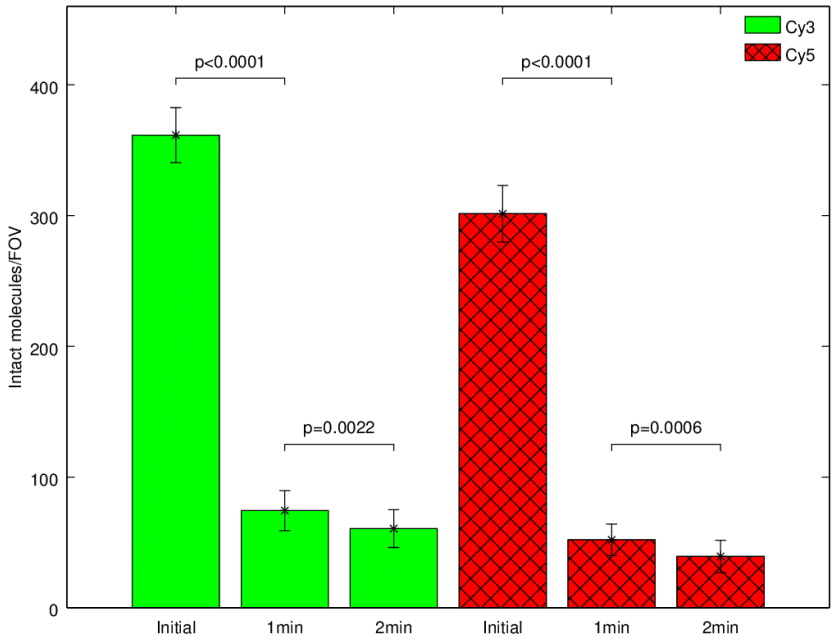

Figure 11: The extent of the digestion is not limited to termini only. We replicated the digestion experiment with internally labeled JJ3-3 constructs spiked in 5nM unlabeled JJ DNA, where the Cy3 (green) and Cy5 (red) are 30 and 24 bases away from the closest termini, respectively. Although the fraction of intact molecules remained slightly higher and reaction speed after 2min was also higher, we still observed a fast digestion trend that takes place within a minute. Sample comparisons were done by two-tailed z-test, error bars show 1 SD.

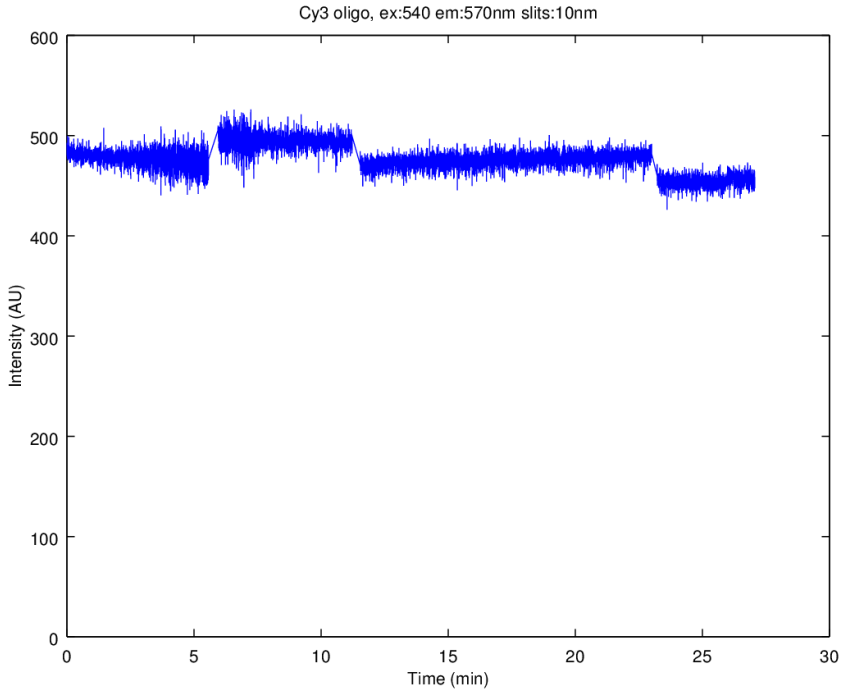

Figure 12: Digestion conditions do not quench Cy3 direct emission. Cy3-derivatized 20mM dsDNA constructs were excited at 540nm and the emission intensity at 570nm was monitored, slit width 10nm in both cases. Cy3 emission is not quenched as individual digestion buffer components are added sequentially to the medium. 0: water only, 6min: 1X NEB4, 12min: 1mM ATP, 23min: gloxy + RecBCD. Final concentrations of the reagents are same as in single molecule experimental conditions.

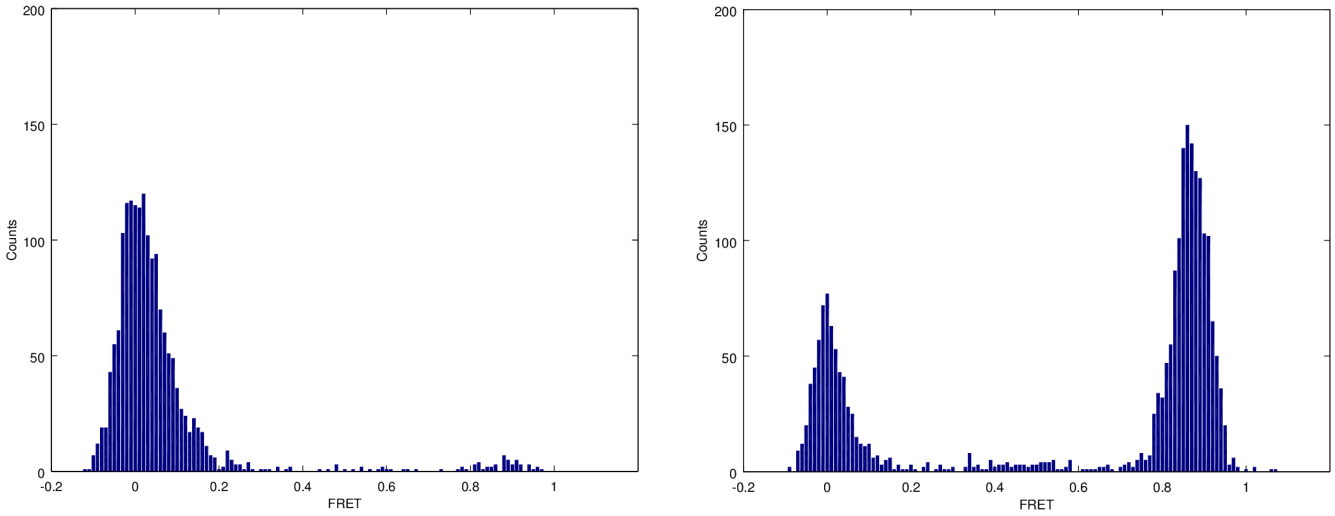

Figure 13: The amplification and purification scheme yields two-state products. Following the PCR protocol described above, obtained sample is essentially fully in low FRET state in low salt conditions (left, 3.6% in the high FRET). Addition of NEB4 induces looping, observed as emergence of the high FRET population which reaches to 67.6% after 1 hour of incubation (right).

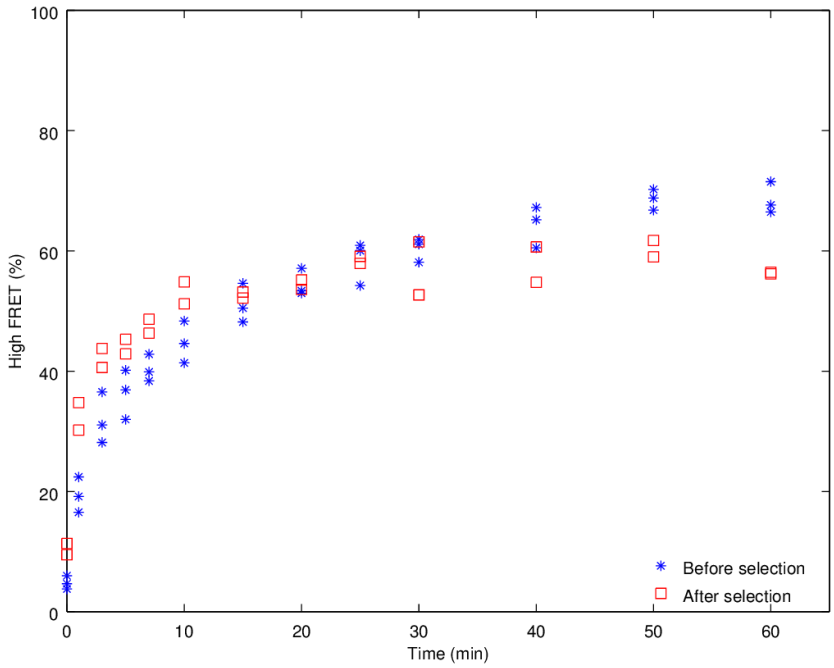

Figure 14: Exonuclease digestion provides contrast. Surface immobilized molecules were subjected to looping conditions in NEB4 for 5 minutes, followed by RecBCD incubation for 2 minutes. The chamber was thoroughly washed with T10 and remaining molecules on the surface were unlooped in 5mM NaCl containing imaging buffer for 60 minutes. Comparison of the looping kinetics of the library after selection with the initial library demonstrates an increased relaxation rate.

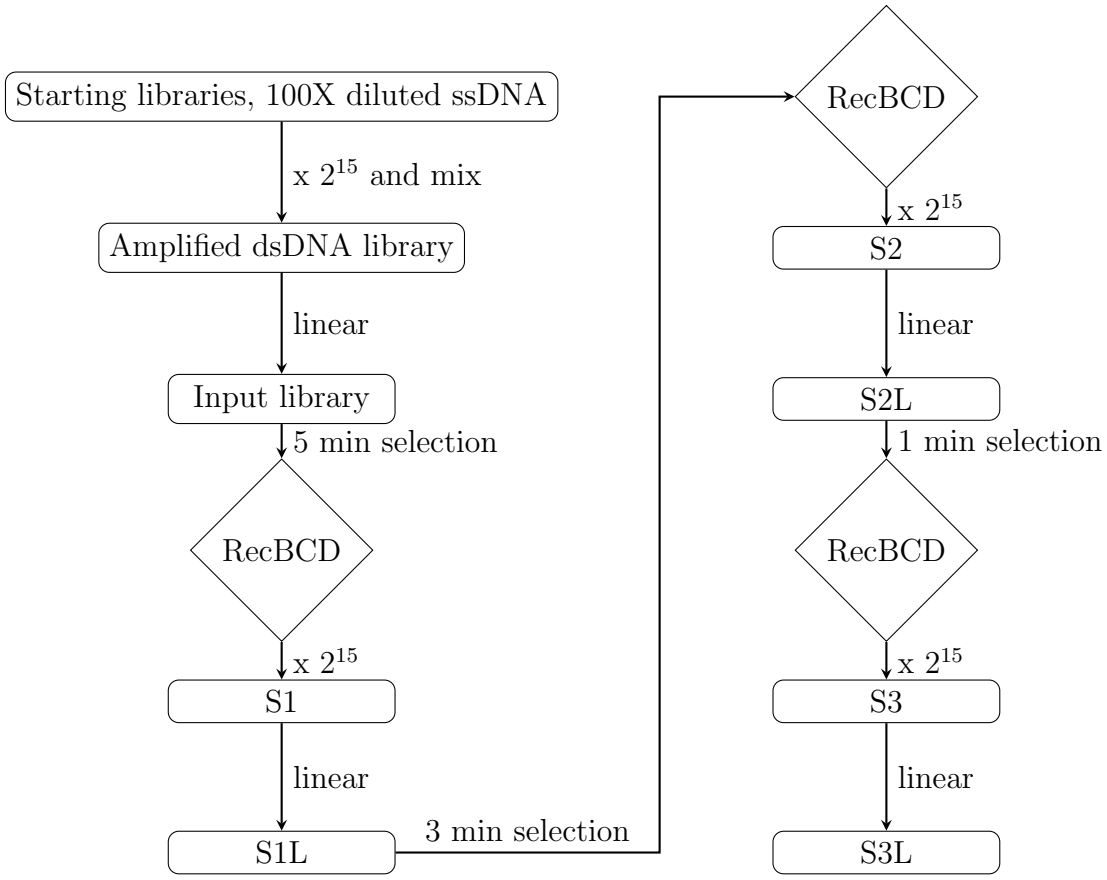

Figure 15: Flowchart of the sequential selection protocol. "linear" refers to two consecutive linear amplifications using modified primers one at a time, whereas exponentials indicate traditional PCR using unmodified primers for exponent numbers of cycles.

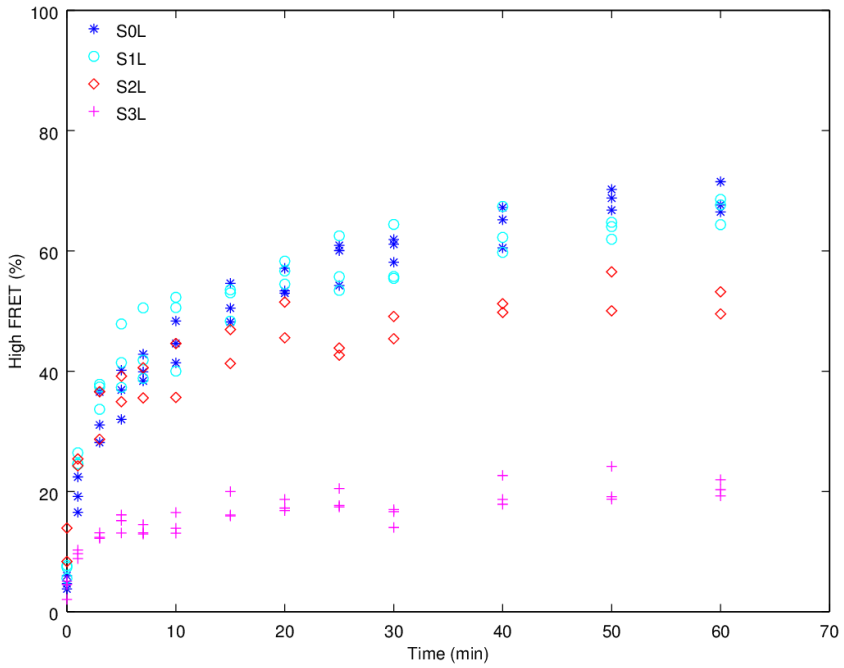

Figure 16: Selection assay renders library stiffer. Looping curves of the library sample after successive rounds of immobilization, selection and re-amplification display a monotonously more rigid character.

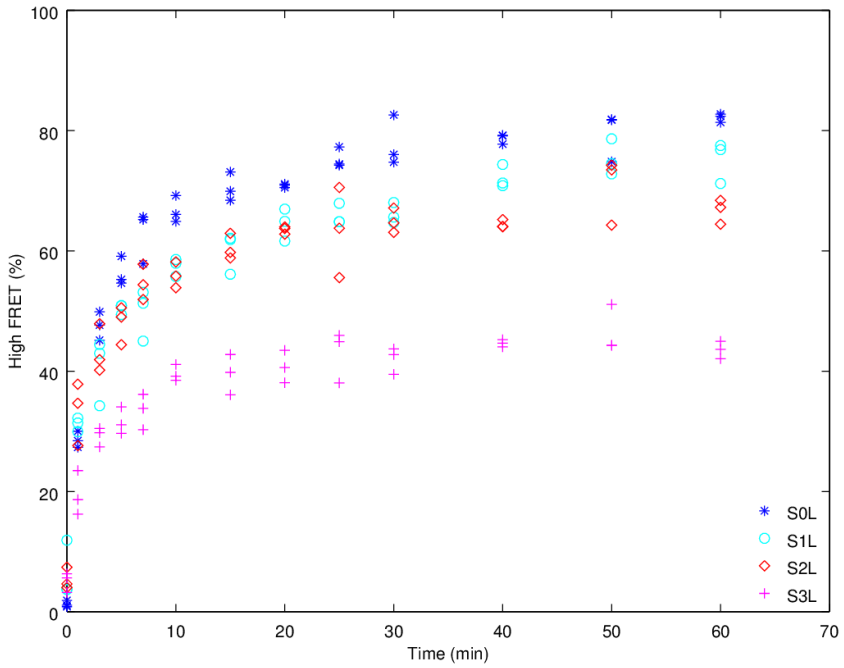

Figure 17: Rigidification of the library not due to artifacts due to choice of buffer. Looping rate measurements for the successively selected library samples were replicated in imaging buffer containing 1M NaCl, which display the same rigidification character observed in NEB4.

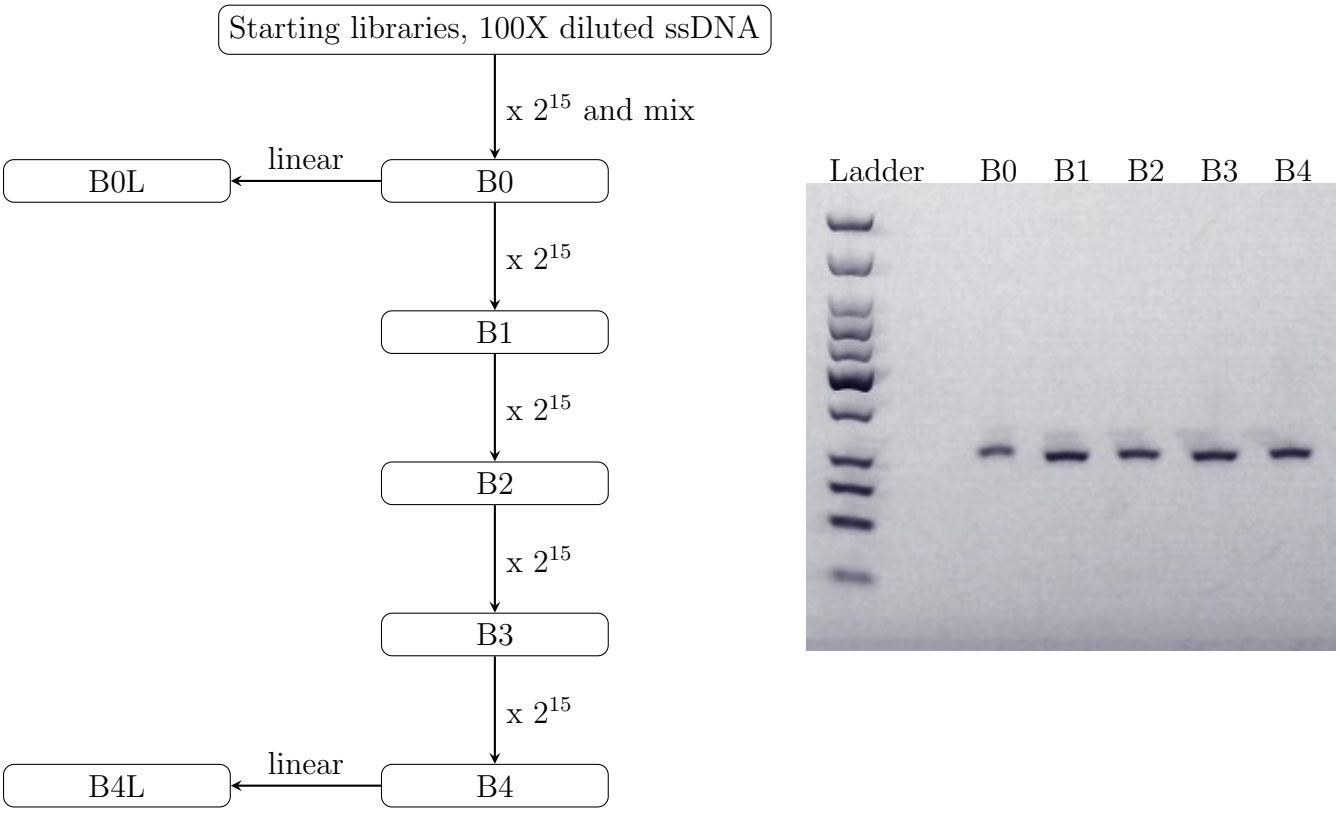

Figure 18: Flowchart of the PCR bias test (left). No extraneous products are detectable after 60 cycles of PCR with unmodified primer pair, nor at any intermediate step (right).

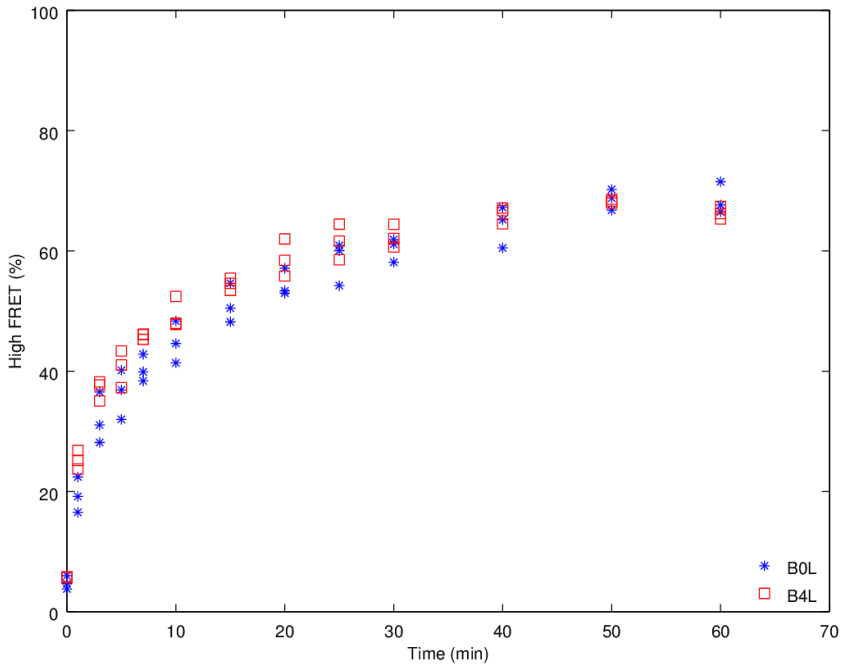

Figure 19: Exponential PCR does not show significant flexibility bias. Comparison of the looping kinetics of the input library and after 60 additional exponential PCR cycles do not reveal a significant difference that could explain the observed gradual stiffening in sequential selection assay.

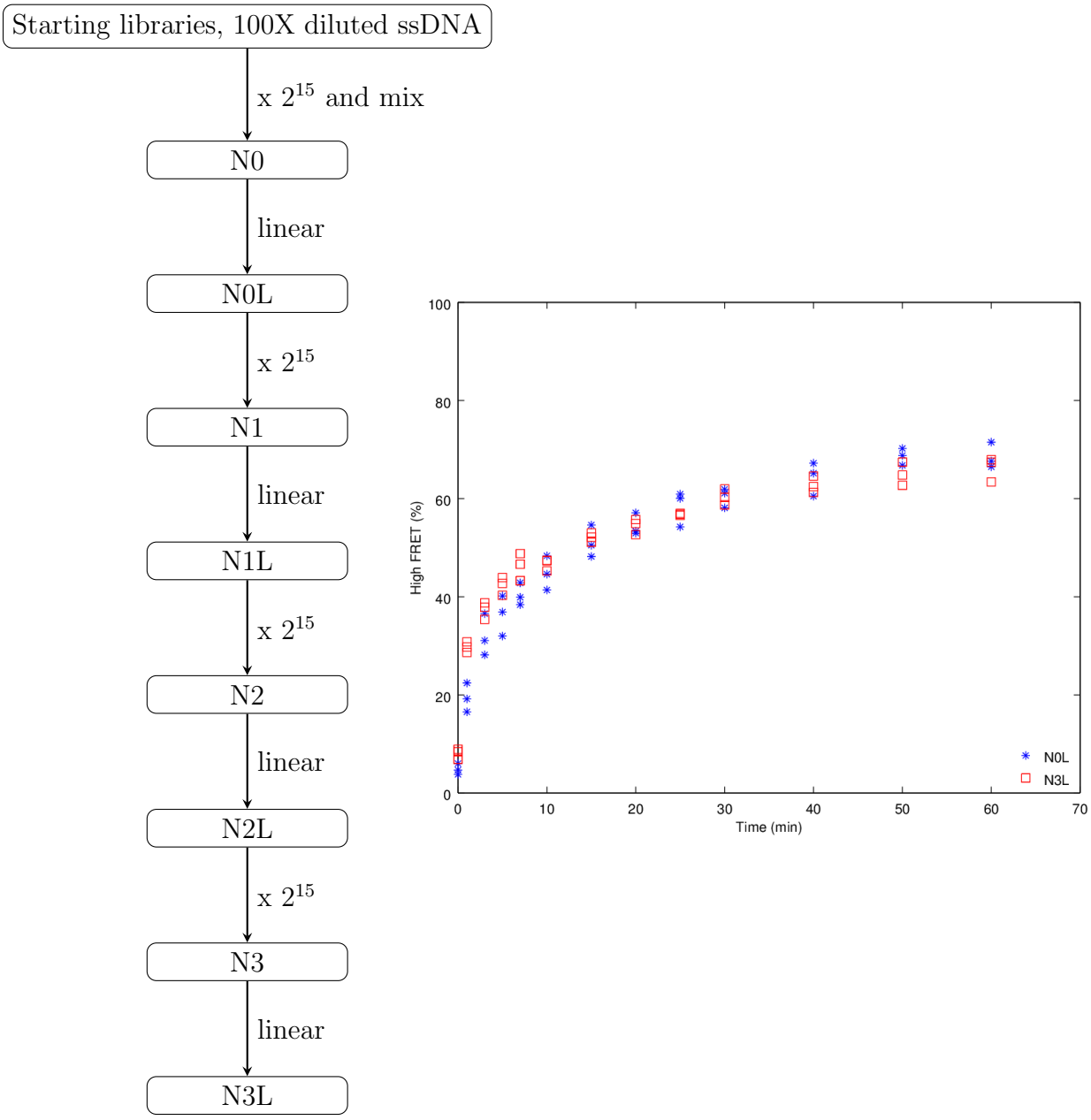

Figure 20: PCR-based modification does not cause accumulation of artifacts. We applied three consecutive construct synthesis using exponential PCR product dsDNA as template (left) and compared looping rates of the initial and final product (right). Absence of a significant difference in looping kinetics confirms that usage of modified primers in our PCR-based sample preparation does not lead to stiffening artifacts that we observed.

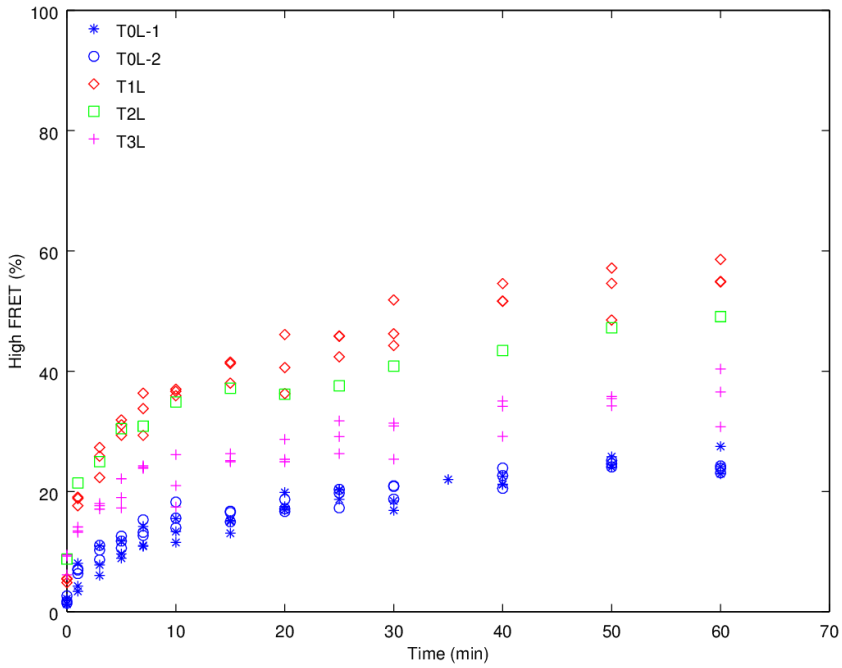

Figure 21: Selection assay does not conserve the characteristic behavior of single molecule specimen. A single species PAGE purified construct obtained from IDT was subjected to the same successive selection protocol described above. Specimen was observed to suddenly become more flexible, which gradually relaxes towards a more rigid state.

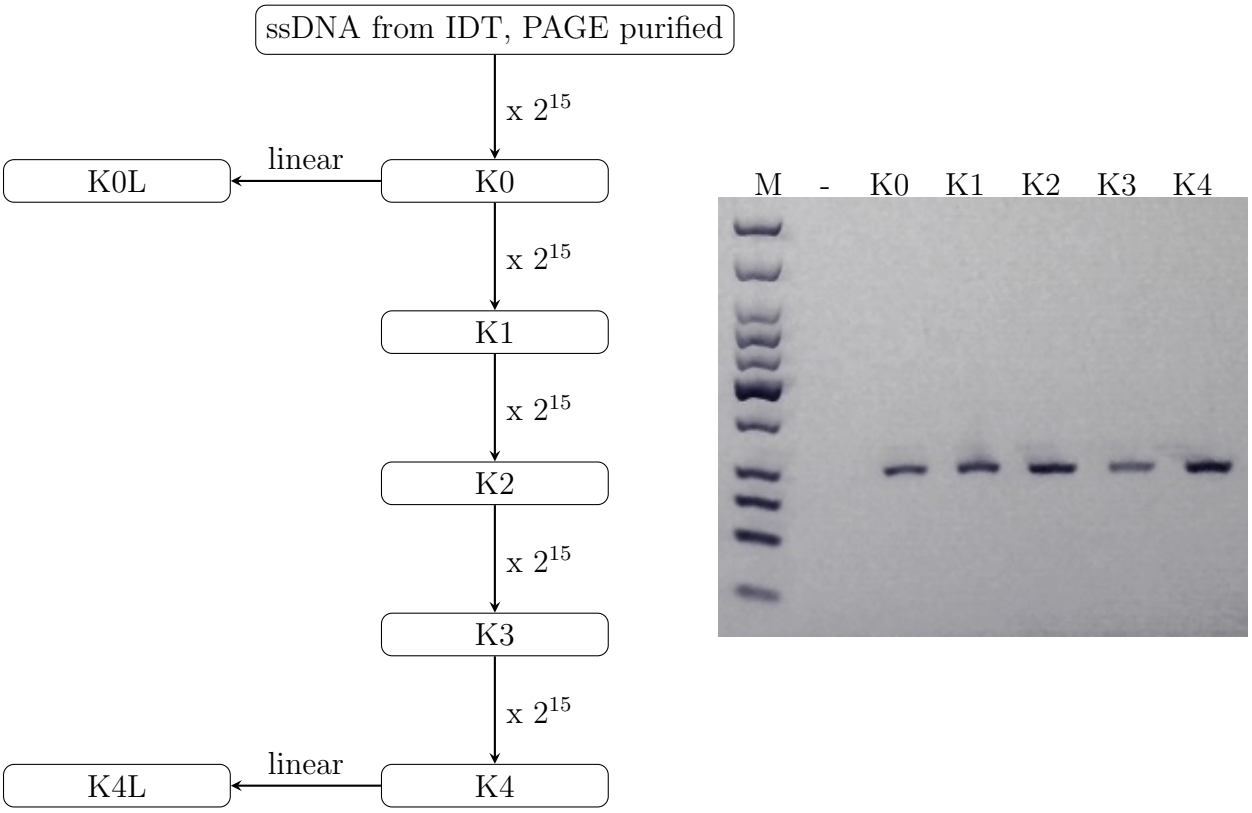

Figure 22: Flowchart of the PCR bias test control by single species (left). No extraneous products were detectable after 60 cycles of PCR with unmodified primer pair, nor at any intermediate step (right). '-' indicates a negative control performed without template against any contamination.

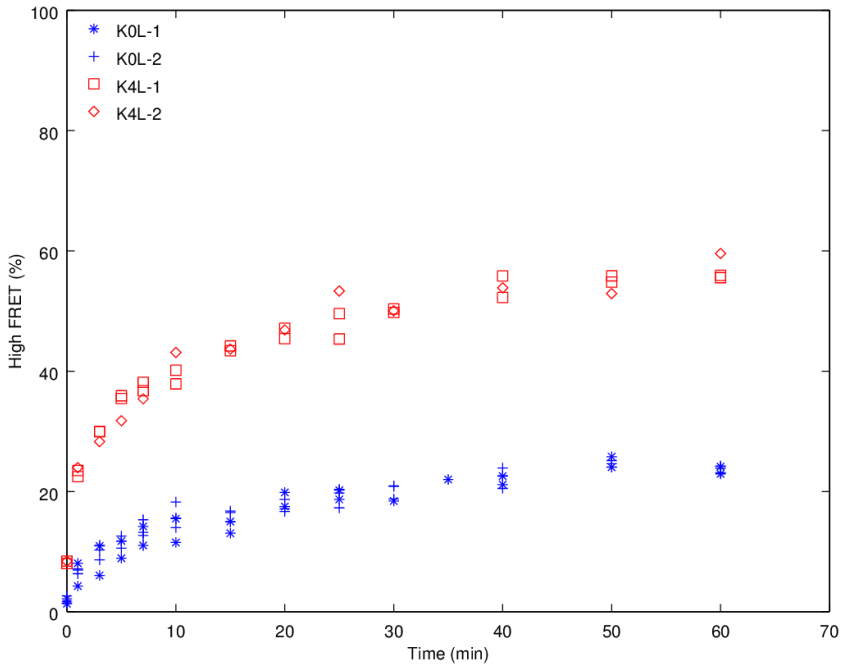

Figure 23: Flexibility of the single species sample increases dramatically after extended numbers of cycles. Indices after dash refer to technical replicates that are based on different sample preparations following the same protocol.

6 References

[1] Strick T., Allemand J.F., Croquette V., Bensimon D. (2000) "Twisting and Stretching Single DNA Molecules" Progress in Biophysics & Molecular Biology, 74:115140.

[2] Wang J., Lu C. (2007) "Single Molecule λ-DNA Stretching Studied by Microfluidics and Single Particle Tracking" J. Applied Physics, 102: 074703.

[3] Vucemilovic-Alagic N. (2013) "Kratky-Porod Model" Accessed on 30 October 2015. http://www-f1.ijs.si/~rudi/sola/KratkyPorodmodel.pdf

[4] Berry R. (2013) IOP: "Biological Molecules 2: Modelling DNA and RNA". Accessed on 31 October 2015. http://biologicalphysics.iop.org/cws/article/lectures/48662

[5] Bouchiat C., Wang M.D., Allemand J.F., Strick T., Block S.M., Croquette V. (1999) "Estimating the Persistence Length of a Worm-Like Chain Molecule from Force-Extension Measurements" Biophysical J., 76:409-413.

[6] Cloutier T.E., Widom J. (2005) "DNA Twisting Flexibility and the Formation of Sharply Looped ProteinDNA Complexes" PNAS, 102(10):3645-3650.

[7] Hagerman P.J. (1988) "Flexibility of DNA" Ann. Rev. Biophys. Biophys. Chem, 17:265-86.

[8] Shimada J., Yamakawa H. (1984) "Ring-Closure Probabilities for Twisted Wormlike Chains: Application to DNA" Macromolecules, 17:689-698.

[9] Mathew-Fenn R.S., Das R., Harbury P.A.B. (2008) "Remeasuring the Double Helix" Science, 322:446-449.

[10] Vafabakhsh R., Ha T. (2012) "Extreme Bendability of DNA Less than 100 Base Pairs Long Revealed by Single-Molecule Cyclization" Science, 337:1097-1101.

[11] Wiggins P.A., Phillips R., Nelson P.C. (2005) "Exact Theory of Kinkable Elastic Polymers" Physical Rev. E. 71.

[12] Xu X., Thio B.J.R., Cao J. (2014) "Correlated Local Bending of a DNA Double Helix and Its Effect on DNA Flexibility in the Sub-Persistence-Length Regime" J. Physical Chemistry Letters, 5:2868-2873.

[13] Wu Y.Y., Bao L., Zhang X., Tan Z.J. (2015) "Flexibility of Short DNA Helices with Finite-Length Effect: from Base Pairs to Tens of Base Pairs" arXiv:1503.03344[physics.bio-ph]

[14] Taranova M., Hirsch A.D., Perkins N.C., Andricicioaei I. (2014) "Role of Microscopic Flexibility in Tightly Curved DNA" J. Physical Chemistry, 118: 1102811036.

[15] Zhang Y., Crothers D.M. (2003) "Statistical Mechanics of Sequence-Dependent Circular DNA and Its Application for DNA Cyclization" Biophysical J., 84:136-153.

[16] Du Q., Smith C., Shiffeldrim N., Vologodskaia M., Vologodskii A. (2005) "Cyclization of Short DNA Fragments and Bending Fluctuations of the Double Helix" PNAS, 102(15): 5397-5402.

[17] Vologodskii A., Du Q., Frank-Kamenetskii M.D. (2013) "Bending of Short DNA Helices" Artificial DNA: PNA & XNA, 4(1):1-3.

[18] Peters J.P., Maher L.J. (2010) "DNA Curvature and Flexibility in Vitro and in Vivo" Quarterly Reviews of Biophysics, 43(1): 23-63.

[19] Rosanio G., Widom J., Uhlenbeck O.C. (2015) "In Vitro Selection of DNAs with an Increased Propensity to Form Small Circles" Biopolymers, 103(6): 303-320.

[20] Law S.M., Bellomy G.R., Schlax P.J., Record M.T. (1993) "In Vivo Thermodynamic Analysis of Repression with and without Looping in lac Constructs: Estimates of Free and Local lac Repressor Concentrations and of Physical Properties of a Region of Supercoiled Plasmid DNA in Vivo" J. Mol. Biology 230(1):161-173.

[21] Schulten K., Kosztin I. (2000) "Lectures in Theoretical Biophysics" Accessed on: 30 October 2015. http://www.ks.uiuc.edu/Services/Class/NSM.pdf

[22] Brogaard K., Xi L., Wang J., Widom J. (2012) "A Map of Nucleosome Positions in Yeast at Base-pair Resolution" Nature, 486:496-501.

[23] Moyle-Heyrman G., Zaichuk T., Xi L., Zhang Q., Uhlenbeck O.C., Holmgren R., Widom J., Wang J.P. (2013) "Chemical Map of Schizosaccharomyces pombe Reveals Species-Specific Features in Nucleosome Positioning" PNAS, 110(50): 201058-20163.

[24] Kestin J., Sokolov M., Wakeham W.A. (1978) "Viscosity of Liquid Water in the Range -8^{o}C to 150^oC" J. Phys. Chem. Ref. Data 7(3):941-948.

[25] Reding F.P. (1958) "The Stiffness Modulus of Polyethylene as a Function of Temperature and Structure" J. Polymer Science 32:487-502.

[26] Gibson, Swoboda, Massey (1964) "Kinetics of Glucose Oxidase" J. Biological Chemistry, 239(11):3927-3934.

[27] Durst R.A., Staples B.R. (1972) "Tris/Tris-HCl: A Standard Buffer for Use in the Physiologic pH Range" Clinical Chemistry, 18(3):206-208.

[28] IDT (2011) "Chemical Synthesis and Purification of Oligonucleotides" Accessed on: 9 Nov 2015, http://www.idtdna.com/pages/docs/technical-reports/chemical-synthesis-of-oligonucleotides.pdf